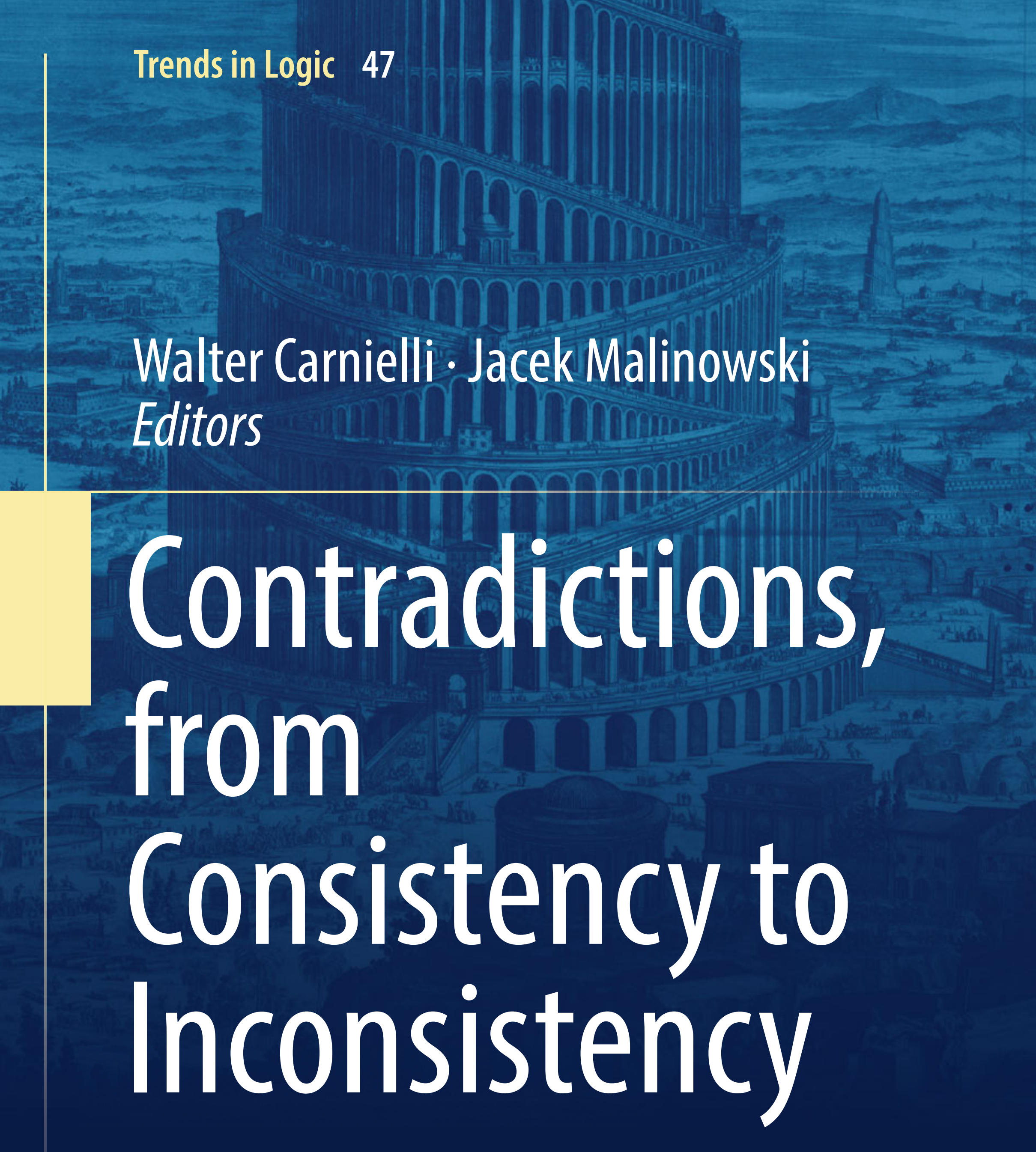


Trends in Logic 47

Walter Carnielli · Jacek Malinowski
Editors

Contradictions, from Consistency to Inconsistency

Springer

Trends in Logic

Volume 47

TRENDS IN LOGIC
Studia Logica Library

VOLUME 47

SCOPE OF THE SERIES

The book series Trends in Logic covers essentially the same areas as the journal Studia Logica, that is, contemporary formal logic and its applications and relations to other disciplines. The series aims at publishing monographs and thematically coherent volumes dealing with important developments in logic and presenting significant contributions to logical research.

The series is open to contributions devoted to topics ranging from algebraic logic, model theory, proof theory, philosophical logic, non-classical logic, and logic in computer science to mathematical linguistics and formal epistemology. However, this list is not exhaustive, moreover, the range of applications, comparisons and sources of inspiration is open and evolves over time.

More information about this series at http://www.springer.com/series/6645

Walter Carnielli · Jacek Malinowski
Editors

Contradictions, from Consistency to Inconsistency

Springer

Editors
Walter Carnielli
Department of Philosophy, Centre for Logic, Epistemology and the History of Science
University of Campinas-Unicamp
Campinas, São Paulo, Brazil

Jacek Malinowski
Institute of Philosophy and Sociology
Polish Academy of Sciences
Warsaw, Poland

ISSN 1572-6126 ISSN 2212-7313 (electronic)
Trends in Logic
ISBN 978-3-319-98796-5 ISBN 978-3-319-98797-2 (eBook)
https://doi.org/10.1007/978-3-319-98797-2

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Contents

Contradictions, from Consistency to Inconsistency

Walter Carnielli and Jacek Malinowski

If something is contradictory, then it is not consistent; but if something is non-contradictory, is it necessarily consistent? If so, there may be nothing between consistency and inconsistency. Thus if we literally apprehend the title of this book, it will be on nothing. However, the title of this book should be understood more broadly. This is because it is not so obvious how we should deal with notions like contradictions, consistency, inconsistency, and triviality. It must not be the case that something is there and is not there at the same time - here is the principle of contradiction in the formulation of Aristotle which, on the one hand, forms the basis of all critical thinking, and on the other hand, it is the object of controversy among the philosophers from Heraclitus through Hegel, to the present day. Jan Łukasiewicz, in his monograph *Aristotle's Principle of Contradiction* (Jan Łukasiewicz *O zasadzie sprzeczności u Arystotelesa*, Polska Akademia Umiejętności, Kraków 1910. English translation after Holger Heine *Jan Łukasiewicz and the Principle of Contradiction*, PhD Thesis, University of Melburne 2013 http://cat.lib.unimelb.edu.au/record=b5152962 access 18-03-2018.) wrote:

> The principle of contradiction is the only weapon against mistakes and lies. If contradictory statements were to be reconcilable with each other, if affirmation were not to nullify denial, but if the one were to be able to meaningfully coexist next to the other, then we would have no means at our disposal to discredit falsity and unmask lies. It is because of this that in every inference in which we apply this principle, for example in an apagogic proof, the concern is to demonstrate the falsity of some statement. And it is also because of this, that the accusation that somebody is caught in contradictions—be it in a scientific treatise or

W. Carnielli (✉)
Centre for Logic, Epistemology and the History of Science - CLE,
University of Campinas-Unicamp, Campinas, Brazil
e-mail: walter.carnielli@cle.unicamp.br
URL: https://www.cle.unicamp.br/prof/carnielli/

J. Malinowski
Institute of Philosophy and Sociology, Polish Academy of Sciences,
Warszawa, Poland
e-mail: Jacek.Malinowski@studialogica.org
URL: http://www.studialogica.org/jacekm/

W. Carnielli and J. Malinowski (eds.), *Contradictions, from Consistency to Inconsistency*, Trends in Logic 47, https://doi.org/10.1007/978-3-319-98797-2_1

> on the witness stand—is such a sensitive matter. Without mercy, it reveals mistakes or lies. Thus, it is the principle of contradiction that makes it possible to victoriously fight a variety of untruths, and on this relies its entire significance.

In the same monograph, Łukasiewicz indicated that the Aristotle's writings included three different meanings of the contradiction principle: ontological, logical, and psychological, although Aristotle himself did not distinguish them. Within the ontological approach 'The same cannot simultaneously belong and not belong to and the same thing'. Within the logical approach 'Contradictory statements are not simultaneously true'. Whilst within the psychological approach, 'Nobody can believe that one and the same thing is and is not'.

The terms *consistency*, *inconsistency* and *contradiction* in the title of this book should be understood in such a broad and multithreaded manner. In the classical logic, based on the bivalence principle, material implication and the modus ponens principle, the Ex falso quodlibet law applies, also known as the principle of explosion. Any material implication with a false predecessor is true. So if the theory allows contradictions, which are false in the traditional logic after all, then any proposition will be its logical consequence. Therefore, if we apply classical logic, any theory containing a contradiction becomes worthless. However, even in this most restrictive comprehension of the contradiction principle, an infraction of the contradiction principle, even only apparent, is an important research tool. As a research tool, we face the contradiction in each apagogical argument (reductio ad absurdum) when in the conditional mood, by assuming the falsity of the argued thesis.

The restriction of the classical logic obtained by rejecting the principle of explosion leads to a wide class of logical systems known as paraconsistent logics. They allow to tackle the local incidental contradictions. If, for example, a vast database contains an error, it is natural to treat it as a contradiction to the fundamental knowledge, external to that database. Applying classical logic to such a database would make it useless since every sentence would be a logical consequence thereof. However, this would be an action contrary to the common sense. After all, it is sufficient to fix the error or at least minimize its effects. Rejection of the principle of explosion makes the database still useful, following the possible isolation of an ambiguous, contradictory fragment. Such locally contradictory but globally useful theories lie just between consistency and inconsistency. However, the subject matter of the papers contained in this volume by far exceeds the scope of the paraconsistent logics.

Not less frequent phenomena are the assemblies of non-contradictory theories which are contradictory to each other. They form a peculiar sort of patchwork of non-contradictory theories. Such theories are not only useful, they also constitute a common phenomenon in science. The relativistic mechanics is contradictory to the classical one; in the first one the velocity of light is constant, while in the second one it depends on the reference system. A photon is a particle or a wave, depending on what phenomenon we desire to explain. The contradiction is not a problem here because the individual theories have different domains. The phenomenon of global contradiction between locally non-contradictory theories, typical of science, shows that in the entirety of its sense, science as a whole lies somewhere between consistency and inconsistency.

What leads us to the contradiction, in science but also in everyday life, are not only the paradoxes, but also conflicting information, bad data and other kinds of uncertainty.

Let us put aside paradox-like sophisms and paralogisms which are based on concealed logical errors. The essence of the proper paradox is that we end up with an unexpected result through a correct reasoning. It can be a contradictory proposition—i.e. an antinomy, or a contradiction in a logical sense—as it is the case for the Russell's Antinomy. However, other paradoxes of infinity lead to the unexpected result that the part does not have to be smaller than the whole. Such a result, even though it is not a contradictory proposition in its own right, contradicts the common-sense intuition and can be treated as a contradiction in the sense of the psychological principle of contradiction. Georg Cantor already in 1899, in his seminal work on the foundations of set theory, referred to multiplicities such that the assumption that all of its elements 'are together' leads to a contradiction, so that it is impossible to conceive of the multiplicity as a unity, as 'one completed thing'. Such multiplicities Cantor called absolutely infinite or inconsistent multiplicities. However, the cases where the totality of the elements of a multiplicity could be thought of without a contradiction as 'being together', or 'forming one thing', would lead to a consistent multiplicity, or a 'set'.

If we take the following property as an axiom: 'A part cannot be equinumerous with the whole', we will easily show that no endless sets exist. And yet this property still seems natural, for many it is obvious, although on the other hand it forms the basis of infinity paradoxes. Which ones of the obvious "truths" we take as an axiomatic starting point for the considerations, is therefore of significance. This fact only became obvious thanks to the discoveries of anomalies related to the concept of infinity. This shows how the set theory, and especially the field of the infinite cardinal numbers, is placed between consistency and inconsistency. The same applies to the selection of axioms and the selection of the primary concepts. A specific selection can determine the contradiction of a theory. The analysis of the theory language and the selection of its fundamental concepts is also an issue which, in fact, lies between non-contradiction and contradiction. This volume consists of 13 papers. Most of them, but not all, are developed around the subtle distinctions between consistency and non-contradiction, as well as among contradiction, inconsistency, and triviality, and concern one of the above mentioned threads of the broadly understood contradiction principle. Some others take a perspective that is not too far away from such themes, but with the freedom to tread new paths.

Jonas R. Becker Arenhart: "The Price of True Contradictions About the World"

The paper examines an argument advanced by Newton C. A. da Costa according to which there may be true contradictions about the concrete world. This is perhaps one of the few arguments advancing this kind of thesis in full generality in the context of a scientifically-oriented philosophy. Roughly put, the argument holds that contradictions in the concrete world may be present where paradoxes require controversial solutions, solutions which in general are radically revisionary on much of the body of the established science. It is argued that the argument may be successfully challenged in the face of the actual practice of science; as a consequence, commitment to true contradictions about the world may be correctly dismissed as unnecessary, at least if the route to contradictions is the one advanced in the argument. Its final part highlights a parallel between da Costa's argument and another typical dialetheist argument by Graham Priest to the effect that paradoxes of self-reference are true contradictions.

Luis Estrada-Gonzalez and Maria del Rosario Martinez-Ordaz: "The Possibility and Fruitfulness of a Debate on the Principle of Non-contradiction"

Five major stances on the problems of the possibility and fruitfulness of a debate on the principle of non-contradiction (PNC) are examined: Detractors, Fierce supporters, Demonstrators, Methodologists and Calm supporters. The paper intends to show what Calm supporters have to say on the other parties wondering about the possibility and fruitfulness of a debate on PNC. The main claim is that one can find all the elements of Calm supporters already in Aristotle's works. In addition, it is argued that the Aristotelian refutative strategy, originally used for dealing with detractors of PNC in Metaphysics, has wider implications for the possibility and fruitfulness of an up-to-date debate on PNC, at least in exhibiting some serious difficulties for the other parties.

Michele Friend and Maria del Rosario Martinez-Ordaz: "Keeping Globally Inconsistent Scientific Theories Locally Consistent"

Most scientific theories are globally inconsistent. 'Chunk and permeate' is a method of rational reconstruction that can be used to separate, and identify, locally consistent chunks of reasoning or explanation. This then allows us to justify reasoning in a globally inconsistent theory. We extend chunk and permeate by adding a visually transparent way of guiding the individuation of chunks and deciding on what information permeates from one chunk to another. The visual representation is in the form of bundle diagrams. The bundle diagrams are then extended to include not only reasoning in the presence of inconsistent information or reasoning in the logical sense of deriving a conclusion from premises, but more generally reasoning in the sense of trying to understand a phenomenon in science. This extends the use of the bundle diagrams in terms of the base space and the fibres. This is then applied to a case in physics, that of understanding binding energies in the nucleus of an atom using together inconsistent models: the liquid drop model and the shell model. Some philosophical conclusions are drawn concerning scientific reasoning, paraconsistent reasoning, the role of logic in science and the unity of science.

Eduardo Barrio, Federico Pailos and Damian Szmuc: "What is a Paraconsistent Logic?"

Paraconsistent logics are logical systems that reject the classical principle, usually dubbed Explosion, that a contradiction implies everything. However, the received view about paraconsistency focuses only on the inferential version of Explosion, which is concerned with formulae, thereby overlooking other possible accounts. This paper proposes to focus, additionally, on a meta-inferential version of Explosion, i.e. one which is concerned with inferences or sequents. In doing so, the paper offers a new characterization of paraconsistency by means of which a logic is paraconsistent if it invalidates either the inferential or the meta-inferential notion of Explosion. The non-triviality of this criterion is shown by discussing a number of logics. On the one hand, logics which validate and invalidate both versions of Explosion, such as classical logic and Asenjo–Priest's 3-valued logic **LP**.

On the other hand, logics which validate one version of Explosion but not the other, such as the substructural logics **TS** and **ST**, introduced by Malinowski and Cobreros, Egré, Ripley and van Rooij, which are obtained via Malinowski's and Frankowski's q- and p-matrices, respectively.

David Gaytán, Itala D'Ottaviano and Raymundo Morado: "Provided You're not Trivial: Adding Defaults and Paraconsistency to a Formal Model of Explanation"

Let us assume that a set of sentences explains a phenomenon within a system of beliefs and rules. Such rules and beliefs may vary and this could have a collateral effect that different sets of sentences may become explanations relative to the new system, while other ones no longer count as such. This paper offers a general formal framework to study this phenomenon. The paper also gives examples of such variations as we replace rules of classical deductive logic with rules more in the spirit of da Costa's paraconsistent calculi, Reiter's default theories, or even a combination of them. This paper generalizes the previous notions of epistemic system. That notion was used to analyze the concept of explanation, using Reiter's default theories and a specific paraconsistent logic of da Costa. The main proposal is a formal framework, **GMD**, based on doxastic systems, which allows the interaction between the theoretical constructs (in this case, explanations), theories and logics to be analysed. The formal framework is intended to be applied to the modeling of scientific explanation, trying, along the way, to shed light on different kinds of interaction between paraconsistency and non-monotonicity.

Bruno Woltzenlogel Paleo: "Para-Disagreement Logics and Their Implementation Through Embedding in Coq and SMT"

Four different disagreement resolution methods were discussed, with special emphasis on the majority voting method. However, it is important to note that para-disagreement logics form a general framework that, in principle, can support other (possibly more sophisticated) disagreement resolution methods as well. The development of para-disagreement logics required a formulation of possible worlds semantics that is technically different from the usual one. Their embedding into the meta-logics of Coq and SMT-solvers also pushed further the state-of-the-art of the embedding approach, as it required the use of arithmetics, which was not necessary in previous work on simpler modal logics. At the same time, the successful (almost full) automation of para-disagreement logical reasoning within Coq and Z3 attests the current level of maturity of these tools even for a domain of application for which they were not originally intended. And indeed, the embeddings described here expand the range of applications of classical interactive and automated theorem provers to the area of paraconsistent reasoning, broadly understood, at least when contradictions are merely apparent as a result of disagreement between clearly identifiable sources. Although the focus here was on propositional para-disagreement logics, this was so just because the propositional level was sufficient to discuss the essence of para-disagreement logics. The embedding into the meta-logic of **SMT**-solvers could be easily extended to quantifier-free first-order logic, and the embedding into the metalogic of Coq can be easily extended to rigid higher-order logic with constant or varying domains (i.e. with actualistic or possibilistic quantifiers). As para-disagreement logics target apparent inconsistencies (e.g. disagreements such as $@_{s_1} P \wedge @_{s_2} \neg P$), they should be regarded as a complement, and not a replacement, to the paraconsistent logics, which handle actual inconsistencies (e.g. $P \wedge \neg P$).

Marco Panza and Mirna Dzamonja: "Asymptotic Quasi-completeness and ZFC"

The axioms ZFC of first order set theory belong to the best and most accepted, if not perfect, foundations used in mathematics. As they imply the axioms of first-order Peano Arithmetic and are presented using a recursively enumerable list of axioms, ZFC axioms are subjects to Gödel's Incompleteness Theorems, and so if they are assumed to be consistent, they are necessarily incomplete. This can be witnessed by various concrete statements, including the celebrated Continuum Hypothesis CH. The independence results about the infinite cardinals are so abundant that it often appears that ZFC can basically prove very little about such cardinals. This paper puts forward a thesis that ZFC is actually very powerful at infinite cardinals, but not at all of them.

Rodrigo A. Freire: "Interpretation and Truth in Set Theory"

The present paper is concerned with the presumed concrete or interpreted character of some axiom systems, notably axiom systems for the usual set theory. A presentation of a concrete axiom system (set theory, for example) is accompanied with a conceptual component which, presumably, delimitates the subject matter of the system. In this paper, concrete axiom systems are understood in terms of a double-layer schema, containing the conceptual component as well as the deductive component, corresponding to the first layer and to the second layer, respectively. The conceptual component is identified with a criterion given by directive principles. Two lists of directive principles for the set theory are given, and the two double-layer pictures of the set theory that emerged from these lists are analyzed. Particular attention is paid to the set-theoretic truth and the fixation of truth-values in each double-layer picture. The semantic commitments of both proposals are also compared, and distinguished from the usual notion of ontological commitment, which does not apply. The presented here approach to the problem of concrete axiom systems can be applied to other mathematical theories with interesting results. The case of elementary arithmetic is mentioned in passing.

Daniele Mundici: "Coherence of the Product Law for Independent Continuous Events"

In his paper *Logische Prinzipien des mathematischen Denkens*, 1905, p.168, Hilbert observed that axioms and definitions in probability theory are a bit confused. As it is well known, the additivity law for the probability of two incompatible events is an axiom, while the product law defines independent events in terms of a preassigned probability function. And yet, independence, just like incompatibility, has a classical probability-free definition in the context of Boolean algebras and propositional logic. By de Finetti's 1932 Dutch Book theorem, the additivity law follows from his notion of a coherent/consistent set of betting odds. The author shows that the product law for (logically) independent events similarly follows from de Finetti's fundamental notion, for Boolean as well as for continuous MV-algebraic events. Thus an axiom and a definition turn out to be corollaries of a consistency notion.

Jose Carlos Magossi and Olivier Rioul: "A Local-Global Principle for the Real Continuum"

This paper discusses the implications of a local-global (or global-limit) principle for proving the basic theorems of real analysis. The aim is to improve the set of available tools in real analysis, where the local-global principle is used as a unifying principle from which the other completeness axioms and several classical theorems are proved in a fairly direct way. As a consequence, the study of the local-global concept can help to establish some better pedagogical approaches for teaching classical analysis.

Marcelo Finger: "Quantitative Logic Reasoning"

This paper examines several similarities among the logic systems that deal simultaneously with deductive and quantitative inference. It is claimed that it is appropriate to call the tasks those systems perform as Quantitative Logic Reasoning. Analogous properties hold throughout that class, for whose members there exists a set of linear algebraic techniques applicable in the study of satisfiability decision problems. The tasks performed by propositional Probabilistic Logic; first-order logic with counting quantifiers over a fragment containing unary and limited binary predicates; and propositional Łukasiewicz Infinitely-valued Probabilistic Logic are regarded from the viewpoint of Quantitative Logic Reasoning.

Walter Carnielli, Hugo Luiz Mariano and Mariana Matulovic: "Reconciling First-Order Logic to Algebra"

This paper propounds an alternative to the traditional cylindric and polyadic algebras for the first-order logic, based on the polynomial representation of first-order sentences. The authors argue that this new algebraic setting can be seen as a legitimate algebraic semantics for the first-order logic, even closer to the primordial forms of algebraization of logic. By employing the notion of M-rings, rings equipped with infinitary operations that can be naturally associated to the first-order structures and each first-order theories, the paper shows that infinitary versions of the Boolean sums and products are able to express algebraically first-order logic from a new perspective. The paper also discusses how generalizations of the method could be lifted successfully to n-valued logics and to the other non-classical logics, helping to reconcile some lost ties between algebra and logic.

Sérgio Marcelino, Carlos Caleiro and Umberto Rivieccio: "Plug and Play Negations"

Within the study of consistency and inconsistency from a mathematical logic point of view, negation is the crucial connective. Negation is used to formulate the principle of explosion known as ex contradictione quodlibet, which is one of the laws that distinguish classical logic from so-called *paraconsistent logics*, that is, calculi designed to model reasoning in the presence of a certain amount of inconsistency. In turn, within logical calculi negation is usually introduced as a derived connective given by the term $\neg p = p \to \bot$, which uses the material implication $\to$, and the *falsum* constant $\bot$. The degree of paraconsistency of a logic, that is its degree of tolerance to inconsistencies, is thus determined by the interaction among these three connectives. This paper considers logics that result from different choices of subsets of the usual inference rules that capture the interaction between implication and *falsum*, thus determining different negations. The techniques used allow for a modular analysis of the logics, providing complete semantics based on (non-deterministic) logical matrices and complexity upper bounds. Using the semantics obtained for the "environment logics", the paper studies the negation-only fragments of each; axiomatizations for each of these negations are provided, and their paraconsistent character is analyzed.

The Conference

The work on this volume began during the Studia Logica conference "Trends in Logic XVI: Consistency, Contradiction, Paraconsistency and Reasoning – 40 years of CLE", held at the State University of Campinas (Unicamp), Brazil, between September 12–15, 2016. The event, centered around the areas of logic, epistemology, philosophy and history of science, celebrated the 40th anniversary of the Centre for Logic, Epistemology and the History of Science (CLE). Some of the papers collected in this volume were presented on the conference. Others were initiated by the discussions that took place. This volume is not so much the aftermath of that conference, but it is more the result of the work that was initiated there.

Acknowledgements At this point, we would like to thank everyone who contributed to the organization of this conference, which was supported by FAPESP and CNPq, and part of the celebrations for the 50 years of Unicamp. We would also like to thank those who contributed to the event and to the development of this volume: the colleagues who served at the Scientific Committee in the persons of Newton da Costa (Brazil), Ítala D'Ottaviano (Brazil), Hannes Leitgeb (Germany), Daniele Mundici (Italy), Heinrich Wansing (Germany), Ryszard Wójcicki (Poland) and Marcelo Coniglio (Brazil), besides ourselves. We also thank the colleagues who served at the Local Committee: Juliana Bueno-Soler (Brazil), Fabio Bertato (Brazil), Rodolfo Ertola (Brazil), Gabriele Pulcini (Brazil), Giorgio Venturi (Brazil), Rafael Testa (Brazil), as well as the staff of CLE and to Jola Monikowska-Zygierewicz who took care about computer layout of the volume.

Special words of gratitude are due to the reviewers. They have significantly influenced the final form of the works contained herein. Thank you wholeheartedly: Bryson Brown, Filippo Casati, Pete Clark, Justin Clarke-Doane, Diana Costa, Janusz Czelakowski, Szymon Frankowski, Rodrigo Freire, Rohan French, Hector Freytes, Ole Hjortland, Juliette Kennedy, Beata Konikowska, Matteo Morganti, Chris Mortensen, Daniele Mundici, Marek Nasieniewski, Ian Pratt-Hartmann, Sam Roberts, Gert Schubring, Sergei Odintsov, Tomasz Połacik, Jean-Michel Salanskis, Samuel Gomes da Silva, Tuomas Tahko, Amanda Vidal, and Anna Zamansky.

The Price of True Contradictions About the World

Jonas R. Becker Arenhart

> *What contradictory beliefs guarantee us, after all, is false beliefs. Contradiction is the short road to falsehood, and if falsehood is not to be avoided, it's not clear what is. In a way, even those who most vociferously urge us to accept contradiction seem to concede this point, for even they reject with horror the prospect of a trivial system in which anything follows. But what is wrong with triviality if not that it assures that even falsehoods will appear as theorems?*
> *If contradiction is to be avoided whenever possible, as surely it is, then proposals that we gracefully embrace contradiction are to be rejected whenever possible as well*
>
> Grim [14, p.27]

Abstract We examine an argument advanced by Newton C. A. da Costa according to which there may be true contradictions about the concrete world. This is perhaps one of the few arguments advancing this kind of thesis in full generality in the context of a scientifically-oriented philosophy. Roughly put, the argument holds that contradictions in the concrete world may be present where paradoxes require controversial solutions, solutions which in general are radically revisionary on much of the body of established science. We argue that the argument may be successfully challenged in the face of the actual practice of science; as a consequence, commitment to true contradictions about the world may be correctly dismissed as unnecessary, at least if the route to contradictions is the one advanced in the argument. We finish by highlighting a parallel between da Costa's argument and another typical dialetheist argument by Graham Priest to the effect that paradoxes of self-reference are true contradictions.

J. R. B. Arenhart (✉)
Federal University of Santa Catarina, Florianópolis, Brazil
e-mail: jonas.becker2@gmail.com

W. Carnielli and J. Malinowski (eds.), *Contradictions, from Consistency to Inconsistency*, Trends in Logic 47, https://doi.org/10.1007/978-3-319-98797-2_2

1 Introduction

Dialetheias are typically defined as true contradictions, that is, formulas of the form α and $\neg\alpha$ both of which are true (see Priest [22, p.1,p.75]). While it is controversial whether there are dialetheias, there is one favorite example of contradiction that is candidate for being a dialetheia: the sentence comprising the Liar paradox (i.e. "this sentence is false"). A feature of the Liar is that it is a *semantic* paradox: it is clear that it does not infect the concrete world, in the sense that it does not concern concrete objects located in space-time. So, even assuming that dialetheias are possible in the actual world, there is no easy example of a contradiction in the *concrete world*. So, is there any prospect for dialetheism to reign in the concrete world?

Doubts about the possible extent of dialetheism are reflected on distinctions such as between *metaphysical contradictions* (contradictions in reality) as opposed to *semantic contradictions* (contradictions featuring "merely" in our models of reality and our theories, but not necessarily representing any features of actual reality).[1] We also find talk about *real contradictions*, meaning contradictions in the concrete world, as opposed to *semiotic contradictions*, i.e. contradictions arising from the workings of language and systematization of knowledge representation (see da Costa [9, chap.3, sec.3]). So, assuming that some contradictions are likely to arise, are they a result of the way the world is, or are they just a sign of our admittedly vague and less than perfect linguistic practices? Answering such questions may provide us a better understanding of the scope and nature of dialetheism (and indirectly, of reality too, of course). Such questions are also of fundamental importance for us to make sense of the very idea of contradictions (a linguistic notion) in reality (which itself is non-linguistic, for sure).

In this paper we focus on contradictions in concrete reality. We stress "concrete" because here we shall not deal with contradictions arising in an abstract realm, such as a Cantorian universe of naïve set theory containing Russell's set, the Newton–Leibniz version of the infinitesimal calculus, and the like. We shall focus our discussion on a fairly neglected argument advanced by da Costa [9, chap.3] to the effect that there may be true contradictions about the real (concrete) world. As far as we know, this is one of the few arguments that attempt to present the claim that the actual world may be contradictory in the context of a scientifically-oriented philosophy. Roughly put, the argument is an attempt to locate contradictions in the world: the contradictory character of reality manifests itself precisely where science meets contradictions in successful theories and where the elimination of such contradictions requires radically revisionary moves, resulting most of the times in artificial amputation of widely held scientific tenets. The heavy price paid in obtaining consistent results acts as a sign that the world may be contradictory, that a contradiction in reality is the main source behind the problem.

---

[1] See also Mares [16] for a corresponding distinction between *semantic dialetheism* and *metaphysical dialetheism*; see also Beall [5] for a distinct terminology but a related attempt at a classification, and also Bobenrieth [8].

Our main claim, however, will be that even though it is not *logically* forbidden to accept such talk about true contradictions, the price of doing so is just too high; there are many difficulties with the strategy advanced by da Costa that seem to point to the fact that consistent solutions are preferable, even at what is viewed as a seemingly exorbitant cost. At least on pragmatic grounds, there are good reasons to prefer to avoid commitments with true contradictions. The main reason, as we shall see, is that accepting true contradictions puts pressure for adjustments on our overall system of knowledge that are at odds with our best current scientific and philosophical practice. Or, put another way: our most successful canons of rationality still seem to privilege consistency; and that happens for good reasons, it seems.

The paper is structured as follows. In Sect. 2 we present, as clearly as possible, the argument leading from paradoxes to (the possibility, at least, of) true contradictions. In Sect. 3 we deal mainly with our objections to the argument. Each objection is presented in a separated subsection. In Sect. 4 we conclude, bringing to light why da Costa's argument should be more well known among friends and foes of dialetheism and paraconsistency: it bears striking resemblances with the main argument by Priest favoring the adoption of dialetheism in the face of the Liar paradox. So, it incarnates the typical strategy to defend dialetheism in many distinct levels.

2 Contradictions in Concrete Reality

To begin our discussions, we shall assume here for the sake of argument that to speak of true contradictions about the concrete world is not a category mistake; that is, the world itself may be the responsible for making contradictory sentences true by containing contradictory facts or something to that purpose. So, we leave aside the worries raised, for instance, by Bobenrieth [8], who has a rather negative view about the possibility of applying the concept of 'contradiction' to the world. In that sense, the thesis to be discussed concerns the claim that there are dialetheias (true contradictions) *about* the concrete world. That being settled, the main problem is that it still remains really difficult to conceive what it would be like to have true contradictions about the concrete world. So, the main question is: what are the prospects for true contradictions about the concrete world?

It seems that even some dialetheists are willing to reject such a radical possibility, confining true contradictions to a purely semantic level (see for instance Beall [6], Mares [16]). Of course, there are arguments for a contradictory world in famous speculative thinkers such as Heraclitus, Hegel, and Marx; more recently, there are attempts to defend that the world is (in some sense) contradictory in association with Eastern religious beliefs. Deguchi, Garfield, and Priest [12, p.371] go on to say that "[i]t is important that samsāra and nirvāna are both distinct and identical at this world". So, by looking at the right places, one may find that claims of a contradictory world are not so rare (see also Priest and Routley [25] for further sources of contradictions in philosophical thought).

However, interesting as those considerations are, we shall confine ourselves to scientifically-oriented arguments about true contradictions, that is, arguments concerning contradictions that arise from our scientific description of the world. Here, da Costa [9, chap.3] (first edition in Portuguese from 1980) offers perhaps the most promising strategy to advance such a point of view from a very general perspective, not based on an examination on a case by case basis (we shall touch on another favorite theme by dialetheists in general, and by Priest in particular, in the final section of this paper). The general idea of da Costa's argument, as we shall see, is that contradictions in the concrete world may be manifesting themselves precisely where no clear agreement over a consistent solution to paradoxical situations *in our best scientific theories* is likely to arise; i.e., troubles in consistently solving a paradoxical situation in empirical science may have their sources in a contradictory reality described by those theories.

To begin with the relevant terminology, following da Costa [9, chap.3] we shall focus here on *paradoxes* involving concrete objects, that is, *prima facie* sound arguments leading to apparently unacceptable conclusions, in particular, to contradictions. Traditionally, contradictions are deemed unacceptable because they are generally thought of as false, a point to which we shall return briefly in the conclusion. When faced with those paradoxical arguments, da Costa [9, chap.3 sec.1] says there are two possibilities:

(i) Those arguments are fallacies (paralogisms), and we may discover and provide for general agreement as to where reasoning went wrong (think about algebraic arguments attempting to prove that $0 = 1$ involving division by zero), or
(ii) The precise source of the problem is a contentious issue. A consistent resolution of the paradox is available only at the cost of substantial theory revision and, perhaps, mutilation of established canons of scientific rationality.

Paradoxes of the second category are called *aporias* by da Costa (see da Costa [9, p.198]). While there are consensual solutions for paralogistic paradoxes,[2] aporias are paradoxes whose "resolutions" are constantly called into question; there is no agreement that an attempted solution succeeded in finding a fallacy in the argument. More than that, the main feature of an aporia is that its consistent resolutions all seem to require revision or mutilations of substantial portions of our currently established canons of scientific rationality.

As an example of a most typical aporia (not about the concrete world yet), consider the Liar paradox. Solutions to the Liar are plenty, and no consensus about which one has got it right seems about to appear. Most consistent solutions to the Liar, such as Russell's and Tarski's, are heavily revisionary of our actual linguistic and scientific practices (see also the discussions of their limitations in Priest [23]; for an overview on semantic paradoxes and some attempted solutions see Horsten [15]). There is no way to treat the Liar as a fallacy without mutilating much of our actual scientific and

[2]Even though they sometimes can be rather sofisticated arguments, from a mathematical point of view.

linguistic practices. That gives the Liar the status of an aporia (see also our discussion in the final section).

Now, the strategy for locating contradictions in the concrete world proceeds as follows. Argumentative mistakes leading to contradictions (paralogisms) are very unlikely to have any source in reality. Their solution is typically fast and, in general, simple, we just have to locate the flaw in the reasoning; they require no revision in the theories in which they are derived. When it comes to dealing with aporias, however, things are much different. Given that heavy revisions are required in order to consistently accommodate an aporia, and that even such movements are controversial and constantly called into question, it seems that they could be anchored in reality. If we could discover general features of aporias, then perhaps by finding candidates to be aporias we would be able to claim that we have at least good reasons to be suspicious that we are facing a contradiction whose source is reality itself. As da Costa [9, p.206] puts it:

> ... if there are real contradictions, there must be differences between solutions to paradoxes that are not aporias reflecting reality, and solutions to other paradoxes, grounded objectively in the real world.
>
> [...]
>
> ... if we are able to detect certain features of aporias properly speaking which clearly distinguish them from paralogisms, we may argue that they probably reflect objective and real contradictions; their resolution, no doubt, will require radical transformations in science. (All translations of da Costa [9] are ours)

So, there is a clear attempt to find a connection between aporias and contradictions in reality. On the one hand (the first implication in the quote), it seems that real contradictions will ground the distinction between aporias and paralogisms. On the other hand (the second implication in the quote), it is advanced that if we can establish a difference between the two kinds of paradoxes, then we may go on and claim that aporias reflect reality somehow. Here we shall focus on the second implication, the one which attempts to establish the existence of true contradictions from aporias. As da Costa indicates, the sign that an aporia reflects a real contradiction is shown in the fact that the resolution of an aporia will require a deep transformation in science (due to the fact that the source of the contradiction is supposedly located in reality). The examples da Costa [9, pp.206-207] furnishes for aporias about the concrete world illustrate this move from aporias (a contradiction in a theory) to contradictions in reality itself. They are the following ones:

(i) *The paradoxical particle and wave nature of quantum entities in the double slit experiment*. As it is well known, in the context of the double-slit experiment quantum entities behave sometimes as particles, sometimes as waves; it depends on how many slits are open (one or two). The most well-known solution to such a paradox, according to da Costa, is the Copenhagen interpretation, which requires severe mutilation of the very canons of scientific reasoning; it confines what may be said about quantum entities only to the context of experimental results, and abolishes the explanatory role of the theory.

(ii) *Zeno's paradoxes of movement and change*. Paradoxes such as the arrow paradox and the Achilles and the Tortoise race are avoided only due to the artificial mathematical machinery of the calculus; our notions of space and time must be substituted by mathematical idealizations (the continuum), idealizations which are very far away from our common notion of space and time. There must be substantial revision of our intuitive concepts in order to overcome the contradictions provided by the paradoxes (i.e. the radical transformation of science due to the solution of an aporia, mentioned by da Costa).

(iii) *The plurality of incompatible interpretations of quantum mechanics*. The Copenhagen interpretation of quantum mechanics is only one of the candidates to interpret quantum theory. Among the alternatives, Bohm, for instance, presented an incompatible interpretation with hidden variables, an account of microphysical reality much closer to classical mechanics. Obviously, many other interpretations incompatible with the Copenhagen proposal and also incompatible among themselves are yet available (many worlds, GRW, modal interpretations, to mention just a few). In the face of many incompatible interpretations of quantum theory, da Costa [9, p.207] asks: couldn't we conjecture that "the difficulties come from our attempts at giving consistent descriptions of an inconsistent reality?"

Before we proceed, a comment on item (iii) is in order. *Prima facie*, this seems to be merely a case of empirical underdetermination, which we could expect to solve in due time: the data do not allow us to choose among the competing interpretations *now*; further investigation may be expected to do so. But why should we consider this as evidence for true contradictions? In order to see the multiplicity of incompatible interpretations as generating an aporia, we take it that da Costa is not only claiming that incompatible interpretations should be seen as accounting for incompatible features of reality, each interpretation getting something right, none of them getting it completely right. Rather, the suggestion is that none of the interpretations is completely problem free: choosing interpretation I_1 leads to difficulties that interpretation I_2 may be seen as solving, and the same could be said of any two consistent interpretations. So, the problem is that it is not possible to provide a unique and consistent account of quantum reality that does not lead us to further difficulties and which does not require radical revision of current science. What underlies the wave-particle duality? What happens in a measurement? What are quantum entities? Interpretations are attempts to answer those questions. However, they do so in incompatible and incomplete ways. Perhaps, and that is how we interpret da Costa's suggestion, the trouble comes from the fact that what underlies the theory is contradictory, so that an understanding of quantum mechanics (as provided by an interpretation) would have to take into account the contradictory reality if it is not to keep missing something. In the end, the trouble seems to be about quantum reality, about what interpretations deal with, not only about the conflict of interpretations.

The key to the argument seems to lie in the fact that a contradictory reality prevents us from a simple consistent solution of aporias, they will always lead to a revisionary transformation. So, the main claims leading us from paradoxes to true contradictions about reality may be summarized as follows: (1) there is a substantial difference between the resolution of aporias and the resolution of paralogisms; aporias require radical revisions ("transformation") in science. (2) This difference may be explained by the fact that aporias are objectively grounded in reality, somehow. (3) If we can identify features of aporias in a paradox, we have evidence for the possibility of contradictions in the world. (4) There seems to be clear cases of aporias in science. (5) So, it is at least possible that some contradictions are objectively grounded in reality.

3 Problems with Real Contradictions

As da Costa [9, p.207] himself notices, the argument is clearly not definitive. There is no a priori way to establish that the concrete world is contradictory; it is a task for empirical science to tell us how the world looks like. In general, it is a fact that empirical science typically eliminates contradictions in the concrete world by changing theory and pursuing consistent solutions. So, what are the prospects for assuming that there may be real contradictions in the concrete world? Our view is that they are not good, and that the above argument is not only inconclusive, but also very unconvincing and problematic.

Of course, it is still open to the friend of contradictions to look for contradictions in other places, for instance, mystic or religious beliefs. However, as we mentioned before, in this paper we shall discuss only the case of using science as a source for true contradictions (and in doing so, we follow da Costa). This restriction poses no serious drawback on our investigation, it seems, given that science seems to provide our most reliable guide to how the world looks like.

The following remarks are conceived to put obstacles on the way of jumping from contradictions in a theory to contradictions in reality, as the above argument attempts to allow us to do. In the final section, we shall discuss a bit further how widespread in the literature on dialetheism such a strategy really is (although it does not usually appear as an argument favoring contradictions in empirical reality).

3.1 The Most Common Objection: Negative Facts

To begin with what sounds as common wisdom against a contradictory actual world, it should be mentioned that one of the most well-known obstacles for true contradictions about the world concerns its alleged commitment with *negative facts*. Suppose the world is contradictory irrespectively of how our concepts describe it. That is, it is the world which makes some contradictory statements (the aporias, perhaps) true. But then, as Priest [24, pp.200-201] remarks, true contradictions in a strong metaphysical sense will require a corresponding ontology: it is required a fact for being the truth-maker of A and (the argument goes) a *negative fact* for being the truth-maker of $\neg A$.

So, it seems, if we want some true contradictions about the concrete world, negative facts will have to be taken into account. It seems that da Costa [9, p.205] goes along similar lines too, when claiming that "[t]he contradiction $A \wedge \neg A$ is real if A and $\neg A$ are true statements, satisfying Tarski's criterion (T) and making reference to *real* states of facts" (emphasis in the original).

At first sight, that application of the T-scheme understood as relating a sentence with a state of facts could well be understood as committing one with negative facts (but see also the discussion on exclusion that follows). If that reading is accepted, notice that one needs not even go as far as to accept true contradictions to be committed to negative facts; the assumption (by da Costa) that a negative sentence corresponds to a fact already seems to require negative facts. That is, commitment to negative facts is not a privilege of dialetheists. Now, although accounting for such negative facts is clearly not impossible, negative facts were looked with suspicious in metaphysics for a long time, to say the least. So, the price of true contradictions, in the metaphysical front, is commitment to negative facts (for further metaphysical exploration of the impossibility of contradictions in reality, see also Tahko [26]).

However, notice that if negative facts are regarded with suspicions by most, so are true contradictions! Dialetheism is by no means the mainstream view about the nature of reality. So, we shall not focus on that kind of objection as our main argument against concrete dialetheias. Although troubles with negative facts are very general and will certainly generate difficulties for any account of true contradictions about reality, it seems to us that da Costa [9] has a logico-philosophical background that may be employed to accommodate such difficulties, even if it does not reassure worries about negative facts. So, as we shall argue, even if negative facts can be accommodated, there are other difficulties lurking around the corner for the friend of dialetheism about the world (for another view on negative facts and prospects about how the negative view on negative facts could change, see Barker and Jago [4]). First, however, let us briefly check how da Costa accommodates facts.

According to da Costa, there is roughly a correspondence between syntactical categories of formal languages employed in the rational reconstruction of scientific theories and ontological categories (the relation between them is the result of an idealized reconstruction through axiomatization, of course, but this is not an issue to be discussed now). For instance, the syntactical categories of individual terms (individual variables and individual constants), predicate constants and variables, and atomic sentences, correspond to the ontological categories of object, relation and property, and fact, respectively (see da Costa [9, p.53] for a detailed exposition). So, when we reconstruct a scientific context encompassing contradictions with the use of an appropriate paraconsistent logic, this reconstruction determines that some propositions of the form A and $\neg A$ will be true, that is, they will correspond to positive and negative facts that must both obtain (at least according to the reconstruction).

So, according to this framework, by adopting a view allowing for true contradictions, the situation may be accommodated by encompassing negative facts too, represented by the negation of a sentence. Obviously, this does not settle the traditional set of problems with negative facts, but it seems that at least da Costa will have a framework in which to deal with the issue, were he to advance a theory about negative facts too. Furthermore, as we said, granted that one is willing to concede that contradictions may arise in the world, the acceptance of negative facts seems just a minor side effect when it comes to accommodate such contradictions. Notice that this move may also work to reassure worries about whether the very idea of contradiction (in general a syntactical or semantical notion) applies to reality (see Bobenrieth [8] for such worries). Anyway, the difficulties with negative facts will have to be discussed, but it seems that in the context of da Costa's approach they can at least be accommodated.

One possible way to avoid discussing negative facts would be to provide an alternative account for negation in terms of material incompatibility (as suggested by Berto in [7], among other places). In a nutshell, negation may be understood in terms of incompatibility as follows: $\neg\alpha$ means, roughly, a state that is incompatible (materially excludes) the state α. Incompatibility or exclusion is a primitive, understood independently of negation. This may avoid negative facts, of course. For instance, 'x is not a wave' may be understood as meaning some state that is materially incompatible with being a wave, as for instance, the state that 'x is a particle', given that particles and waves exclude each other. No need for negative facts such as 'x is not a particle'. Also, this is said to grant a metaphysical version of the law of non-contradiction: one cannot have incompatible states obtaining together, given the very meaning of incompatibility.

Notice that incompatibility is compatible with da Costa's framework on the relation between logic and ontology. In fact, all that was required there was that atomic sentences correspond to facts, but nothing was said about sentences composed by the use of connectives such as negation, for instance. Perhaps, by making some adjustments, with the use of this proposal of incompatibility it would be possible to keep both the idea that a true sentence of the form $\neg\alpha$ corresponds to a fact, and that it does not need to correspond to a negative fact. So, another approach to true contradictions could be envisaged that avoided the main traps against negative facts.

However, that suggestion really does not seem to help us with the debate on there being true contradictions. Incompatible states, by definition, cannot obtain together, there are no true contradictions when negation is taken to represent a form of material incompatibility. Of course, this was the main motivation for Berto [7] to advance the proposal to begin with: the idea was that there is a sense of contradiction that even dialetheists should accept. As a result, whenever two states may obtain together, they are not incompatible, and the conjunction of statements referring to them does not result contradictory. Let us keep with our example of particles and waves in quantum mechanics. Here, da Costa's suggestion that particles and waves in quantum mechanics may be a case of a true contradiction only means that being a particle could, in fact, be compatible with being a wave. But then, given the suggested understanding of negation, that would no longer be a case of contradiction: 'being a particle' would not imply 'not being a wave', and 'being a wave' would not imply 'not being a particle'. That is, every time a contradiction (understood in terms of incompatibility) seems to be true, the very contradiction disappears, given that the incompatibility disappears. True contradictions are no longer possible. As a result, this suggestion may avoid negative facts, but it does not allow the debate to begin with. So, we shall not pursue this suggestion here.

Instead of insisting on those typically metaphysical problems, our worries are focused on the relation between the true contradictions and their alleged source in science. We shall concede the argument without putting emphasis on the metaphysics of negative facts, or on banning true contradictions with a metaphysical primitive such as incompatibility.

3.2 Dictating Science from the Armchair?

So, by accepting the argument from paradoxes to true contradictions, what are we conceding, in fact? In this section, we shall discuss a methodological objection to accepting true contradictions in our scientific theories, at least on what concerns concrete objects. The general idea is as follows: the argument from aporias to true contradictions would seem to imply that science could be generally mistaken in eliminating aporias when it indeed does so. In other words, if the argument is reasonable and there are (or could have been) at least some real contradictions behind the aporias, then, it seems, their elimination is not warranted! To eliminate a contradiction would put us in the wrong path, we would be deviating from truth.

Before we proceed, we should make it clear that we are assuming that the discussion on true contradictions requires a form of scientific realism. In fact, in order to discuss whether our current theories provide for real contradictions, it is required that we take those theories as correctly (truly) describing at least some aspects of reality. That is most important when we notice that discussion about true contradictions concern, most of the times, unobservable posits of scientific theories (as we shall emphasize later).

Now, that realist stance is almost always coupled with a kind of naturalistic approach to philosophical methodology: philosophy should not dictate the development of science. In another slogan: scientific practice has precedence over philosophical speculation. If philosophy and the march of science conflict, it seems, it is the first that must give precedence to the second. In the face of such commitments to realism and naturalism, the idea that some theories may have provided for true contradictions seems largely unwarranted.

Let us focus, for the time being, on past cases of aporias that were eliminated in our current science. On the one hand, we may rest content with the eliminations of contradictions made by science so far and keep with the development of science or, on the other hand, we may hold on to the argument that aporias are signs of contradictions and claim that even though science has indeed eliminated some contradictions of our description of reality, they are still there, due to their character of aporia; we have got the wrong track with such an elimination. In the first case, there are no more worries with real contradictions, and there is nothing else to be said; the argument is abandoned, at least for those past cases. In the second case, it is implied that science (through the collective action of scientists) has made some wrong moves in providing for such an elimination, so that the contradictions should be allowed to stay. The fact that a new theory does not describe those contradictions does nothing to remove them from reality, of course! The main problem with this move is that the defense of true contradictions ends up being made on purely speculative grounds, given that it is no longer grounded in actual science (given that scientific evidence for them was eliminated in the course of science). So, to keep with some contradictions requires that we see philosophy (or some philosophers) as having a privileged take on how science is to be developed. Philosophers (the dialetheists, only) would end up legislating for science. In other words, the idea is that as soon as a contradiction is eliminated from a scientific theory by a new development, the argument does not deal anymore with current science.

The difficulties in this case could be softened by the adoption of more moderate views on the relation between metaphysics and science. We could assume that the exploration of ontological possibilities is also a task for the philosopher (see specially da Costa [10, p.284], Morganti and Tahko [18]). So, the possibility that the world is inconsistent is to be taken seriously, at least as a metaphysical possibility. What makes this option scientifically relevant, according to this view, is that at least some scientific theory could be seen as providing indirect evidence for the actual truth (or approximate truth) of this inconsistent option. Ontological possibilities can be discarded only after they have proved to be clearly incompatible with current science or when they are clearly deficient in comparison with other candidates available (see also Arenhart [2] for such methodological discussions on ontology).

However, this moderate view does not provide for any comfort for the dialetheist. While the strategy is interesting and plausible in general as part of metaphysical methodology, if it is to be taken as relevant, the idea that a contradictory ontology is possible will still have to be a legitimate possibility for an actual scientific theory. As we have seen, science typically eliminates its contradictions, so, given that we are using science as a test for the very plausibility of ontologies, the fact that there is a possibility involving contradictions would not work. The fact is that science, at least as it has developed so far, ends up eliminating contradictions, and the contradictory possibilities are typically discarded in the way to scientific change.

The alternative, then, would seem to be to claim more radically that contradictions *should not* be eliminated, neither in past theories, nor in future theories. However, by doing so one would end up putting philosophers much above scientists on what concerns scientific development. This would amount to giving up the idea that philosophers should not dictate science from the armchair. In this sense, it would be hubristic to hold to the "true contradiction" option on what concerns our current scientific theories, to say the least. That is, accepting true contradictions just because it is an available option, it seems, is not the best option. One would end up hanging to the science that could have been, instead of the science that actually is (Priest [23] seems to do something like that in relation to set theory).

Of course, to hold that no contradiction will ever arise in a theory that could come out true is also to legislate for the future. Science could well end up choosing dialetheism some time, so that cases of contradictions must be evaluated on a case by case basis. To keep open the possibility of a true contradiction in future theories, da Costa [9, p.208] remarked that it is easier to prove that contradictions are real than that they are not: one single case could prove the positive thesis, while it is difficult to prove non-existence of something, except on speculative (non-naturalistic) grounds. However, here the table may be turned on da Costa: one is never certain that the existence of an aporia really refers to a true contradiction; given that it seems rather unlikely that scientific methodology will change on what concerns elimination of contradictions, it is much more likely that any contradictory theory is adopted temporarily, and that the next theory will eliminate (or try to eliminate) the aporia. There is no *definitive evidence* that could be furnished in order to establish that a contradiction is indeed true.

There are further reasons, it seems, for us to think that the argument according to which aporias may be pointing to true contradictions is problematic. Let us check.

3.3 Unequal Treatment of Observable and Unobservable Contradictory Objects

A rather different problem is that the argument for true contradictions is at odds with another resolution by da Costa and even other dialetheists such as Beall [6] on what concerns concrete objects. The difficulty is related to the asymmetry on the treatment of aporias and paralogisms about concrete objects. To illustrate the issue, consider the famous paradox of the catalog. If there were a catalog C of all the catalogs of a library that do not mention themselves, we would end up with a situation clearly leading to a contradiction: C mentions itself if and only if it does not mention itself. For such paradoxes, there is a fast and clear diagnosis: by *reductio*, it is sensible to blame the assumption of the existence of C as the source of the contradiction. Classical logic applies to concrete objects as catalogs (this is da Costa's own diagnosis, see [9, pp.199-200, p.205]; Beall [6, p.6 and p.16], on the other hand, does explicitly claim that classical logic applies to sentences referring to the world in general, not only catalogs).

However, aren't particles and waves in quantum mechanics just as concrete as catalogs? Obviously, quantum entities are not observable, but it seems that they are just as real if we assume a form of scientific realism (which, again, seems to be required if we are to deal with true contradictions arising from them and have the T-scheme interpreted as suggested by da Costa, in the previous section). But then, when it comes to dealing with contradictions, why are the treatments different for catalogs and for quantum entities? Why do we conclude that by the force of the contradiction the catalog C does not exist, while quantum entities exist and are (possibly) contradictory? There is an awkward asymmetry in the treatment of both cases which is left unexplained. Shouldn't similar cases preferably have similar treatments?

Obviously, one may claim that the catalog case is a paralogism, while the quantum case is (at least *prima facie*) a legitimate aporia. But justifying the application of that distinction in the case of concrete objects is precisely the point in question. The cases may not be different, for all we know. As da Costa has acknowledged, science eliminates its contradictions, so we may end up discovering that the quantum case (if it really presents a contradiction, of course) is just a paralogism too, but one of a much more difficult resolution, given all that is involved. In this sense, the argument for true contradictions proves too little: it is not enough merely to affirm a distinct treatment for quantum entities and for catalogs. By assuming the resolution that contradictory catalogs do not exist, there is nothing to prevent us from claiming that there is a possible resolution of the paradox in quantum mechanics which proves, for instance, that no entity is both particle and wave at the same time, just as in the case of the catalog. In fact, a more developed theory may provide for another kind of more fundamental entities (maybe the standard model or even string theory?) that allow for refinements that may lead to a resolution of the issue. It just happens that such a solution requires much more efforts than the case of the catalog, but both kinds of entities would be in the same category.

The plausibility of the argument, then, seems to rest on the distinction between observable and unobservable entities (problematic as it is): contradictions for observable entities are (or at least ought to be) paralogisms, contradictions for unobservable entities may well be aporias. The only reason for the distinct treatment of those entities comes from the difficulty in dealing with quantum particles, while it is relatively easier to deal with catalogs. That, however, has less to do with the nature of the problem than with the nature of the entities involved. While catalogs are simple medium-sized artifacts, quantum entities are complex and highly removed from our intuition, described by a complex mathematical theory, so no wonder a more difficult solution to problems concerning them is likely to be required. In this case, the complexity of the answer and the kind of revision it requires is not to be taken as a sign that a contradiction is involved, but rather it concerns the very complexity of the subject-matter under study.

So, if one is not willing to concede that paradoxes for unobservable entities may be paralogisms too, it seems plausible to require that one should be forced to accept that some paradoxes for observable entities may be aporias too, on the grounds of having the distinction unmotivated. In this case, however, the argument inconveniently proves too much: there could be contradictory medium-sized objects such as catalogs. But that is something we would hardly assume. So, assuming that some unobservable entities are responsible for true contradictions, while observable ones are not, may sound tricky: it is the unobservable nature of the entities which prevents reasoning about them from being treated as paralogism. Unobservable entities, at least in the case explored, involve such difficult problems that observable entities do not seem to engender. The idea of true contradictions benefits from unobservability to treat similar cases as distinct, a treatment that, as we have seen, is not justified from the point of view being discussed. In the end, it seems that one should either accept that contradictory catalogs may also exist or else assume that quantum entities are not contradictory, with the progress of science required for the paradox to be solved.

Of course, the relation between unobservability and the complexity of the problem is contingent on the case under study, but it illustrates a pattern that may appear in many cases. Unobservable entities may engender much more difficult problems, so no wonder their solution will cause us much more trouble. This leads us to another difficulty posed by the relation between aporias and contradictions, as described by da Costa's argument: that the elimination of aporias involves always a kind of negative move, mutilating the whole body of science available at a given time.

3.4 The Solution of Aporias and Mutilation of Reason

Suppose we admit that some contradictions are true, and that their source is concrete reality. Then, the use of a system of paraconsistent logic seems to be warranted under this assumption to tame the contradiction. But then, granted that the only problem with the contradiction was that a classical framework has an explosive consequence relation when contradictions are present (i.e. from A and $\neg A$ any formula B whatsoever follows), after a paraconsistent logic is assumed nothing else is required on what concerns those entities. That is, to keep with the case of quantum mechanics, if entities are really particles and waves, and that's the end of it, there is no requirement for further investigation, at least on what concerns the source of the original problem. So, if aporias were here to stay and did in fact represent real contradictions, then aporias would be easily accommodated by the adoption of some paraconsistent logic, leaving the rest of science untouched.

But that is not how science works: the quest for understanding seems to require that we search for a solution to the aporia; in the case of particles and waves in quantum theory, for instance, by the construction of quantum field theory or even deeper theories that are still being sought. So, the diagnosis that a contradiction is the source for an aporia may be just misguided: in fact, it is a sign that we got the wrong path and it is time for another scientific revolution. That is how typically we deal with a contradiction: it is a sign for us to retrace our premises and to revise our beliefs. Keeping with contradictions by the adoption of a paraconsistent logic would be detrimental to the evolution of science. In this sense, keeping the contradiction would amount to a bigger mutilation of science: that contradictions be accommodated by theoretical advances in the underlying logic, not of the specific branch of science under consideration.

Consider for instance the actual situation in fundamental physics, where quantum mechanics and general relativity are inconsistent (or, perhaps more clearly put, incompatible). Attempts to provide for a fundamental theory by unifying the two have faced many difficulties, perhaps most of the difficulties that a typical resolution of an aporia would engender. But is it a case of an aporia? Maybe a paraconsistent logic could be employed in contexts requiring the use of both theories and we could claim that it is really a case of inconsistency in the world. But then, science would be left just as it is, except for a change of logic. The search for unification in physics would be pointless. Certainly, assuming true contradictions about the world in this case would not help us in advancing science (for further case studies in which progress in science was achieved by the elimination of contradictions, see Mosterín [20]).

Other things being equal, keeping the contradictions is by far not the most fruitful option. As Priest [21, pp.421-422] has argued, one should balance between theoretical virtues in order to choose the most plausible theoretical hypothesis. When a contradictory option is the most fruitful one, this should not deter us from choosing it. But, as it seems, at least in the case of empirical sciences, the paraconsistent option has only non-triviality to offer. Indeed, choosing to keep contradictions and embrace a paraconsistent logic adds no new explanatory power, simplicity or unification (more on these in the next subsection). So, the path to the elimination of contradictions is the one that always seems more promising, with the success of science attesting it.

That may shed some light on the claim that the resolution of an aporia requires substantial changes in science. That science is required to change in order to accommodate recalcitrant experiences should not be seen in such negative lights. In fact, that is just how science manages to solve its problems. Some problems are such that they cannot be accommodated in our current theories, so a new theory is created. Substantial change in our intuitive frame of the world should not be seen as necessarily negative; in fact, that is what twentieth century physics brought us most forcefully ever since relativity theory and quantum mechanics appeared. So, if the elimination of aporias requires that some intuitions should go, so be it. Keeping with contradictions instead of fomenting their resolution invites a conservatism that is not the mark of science so far.

Perhaps the resolution of Zeno's paradoxes through the tools of the calculus is a better illustration of the price paid to remove contradictions? True, it introduces counterintuitive tools and idealizations that take us far from our comfortable intuition of space and time. However, that is precisely what science is about most of the time. There is no easy description of reality without such idealizations and abstractions, they are part of science and cannot be eliminated. Furthermore, their resolution is currently leading to further investigation of discrete consistent solutions to the paradoxes (see Ardourel [1]). That is, even if the claim that the substitution of intuitive space and time by the continuum were sound, it would provide no ground for us to prefer to keep with the aporias.

Still concerning the negative press given to the mutilations required by the resolution of aporias, consider the example of the Copenhagen interpretation. This is certainly not a good case against the removal of aporias. In fact, the Copenhagen interpretation, although it is certainly not a cohesive and unified view of a group of physicists, indeed had such a strict view of science and its goals (considering Bohr as a main representative of the Copenhagen interpretation). But that is not a feature of the elimination of aporias in general. Rather it is a peculiarity of the Copenhagen interpretation! More: other interpretations, such as Bohm's, attempt precisely to solve the difficulties by keeping most of the traditional ideas concerning the description of a physical system. So, the Copenhagen interpretation does not count as a typical move in the resolution of aporias. That is, restriction to empirical data available in experiments is a mark of the Copenhagen interpretation, but of this interpretation only, not of the elimination of contradictions in general.

One worry with those arguments may be as follows: it may well be the case that our ability to eliminate aporias is not matched by reality, which is contradictory. Our rational desire to keep consistency does not match reality itself. That suggestion may only be pursued on a complete dissociation of our epistemology (here, incarnated in our best scientific theories) and metaphysics. However, that is incompatible with the realist view adopted here. If a contradiction is to be present without it being suggested by science, then, it will have to be suggested by alternative means. In that case, of course, the realist will prefer to stick with science instead of other incompatible sources of knowledge of reality, as we have already argued.

3.5 Contradictions and Theoretical Virtues

Let us focus once again on the negative view of change in science that is allegedly brought by the resolution of aporias and consider what a paraconsistent dialetheic alternative would bring. Recall that it was said that the resolution of aporias requires deep revision of scientific canons of rationality, and that such a revision somehow does not always work to the benefit of our understanding. For instance, it was advanced that the contradiction between wave and particle in quantum mechanics could be solved by the Copenhagen interpretation, which on its turn mutilates science, banning legitimate search for explanations and requiring the abandonment of other traditional features of science. But again, that is a peculiar feature of the Copenhagen interpretation, a feature that is not common to all of the consistent solutions to aporias in general (far from that). So, the example is favorable to the dialetheist, but it is not a typical case of scientific development by the resolution of problems.

Of course, there is no easy and consensual consistent solution to wave-particle duality yet, although there are plenty of attempts. The same could be said about quantum superposition and entanglement (issues an interpretation has to deal with). As da Costa and de Ronde [11, sec.3] have stressed, perhaps it is time for us to abandon consistent attempts to understand the working of superposition and try to face the fact that a contradiction is involved. Also, as Priest [21, pp.421-422] has remarked, consistency is not the sole theoretical virtue involved in theory choice: when an inconsistent hypothesis delivers unification, explanatory power, elegance, and so on, it should be considered as legitimate. Maybe it is time to face the fact that quantum reality is inconsistent.

But let us not go so fast. Let us consider what the dialetheic solution to a paradox would give us in terms of theoretical virtues. Assume that we are employing a paraconsistent logic in order to accommodate contradictory particle-wave behavior, or superpositions and entanglement (perhaps a paraconsistent set theory, as suggested in da Costa and de Ronde [11]). Are we able to save the so-called traditional features of science (whatever they are taken to be) that are said to be eliminated by consistent solutions? It seems that we are not, at least not by merely employing a paraconsistent logic. By embracing a contradiction we by no means assure that explanatory power and other virtues are safeguarded or improved. What are the explanatory powers of contradictory objects? The same as the consistent ones, it seems. If a paraconsistent quantum theory is more explanatory than a consistent rival, then there are means to distinguish the theories, it seems, and we may judge their empirical adequacy. However, so far no such case has emerged.

That is, to the best of our knowledge, no one has managed to provide for a paraconsistent formulation of quantum mechanics. Our claim is that such a formulation, if it is going to appear, will have to preserve the empirical results of standard quantum theory. But if the theory will be formulated with the mere adjunction that superpositions are contradictory, then that will not enhance our understanding of what is going on; rather, the other way around: it leaves us even more distant from a clear understanding of the situation. It is enough to recall popular expositions of quantum mechanics that attempt to convey to the general public a flavor of quantum superposition by holding that quantum particles may be at two places at the same time, or that Schrdinger's cat may be both dead and alive. Those claims certainly add to the mystery of the theory, but shed no light on what is going on (see the discussions in Arenhart and Krause [3]). To know that they are contradictory helps nothing with the quest for understanding.

Perhaps a case could be made by the paraconsistent *interpretation* of superposition, as advanced by da Costa and de Ronde [11], and de Ronde [13]. The difference is that the contradiction comes now at the level of interpretation, not of the formalism (with interpretations thought of as adding another level of theory to the formalism, now at an informal level). According to this view, superpositions are sometimes real contradictions, even though this is not present at the formalism. Consider for instance an electron in a superposition of states between spin up and spin down. Potentially, the electron is both spin up and down, so a contradiction obtains. Could that introduce explanatory power to quantum mechanics or restore some of the traditional virtues of science that are typically abandoned in quantum mechanics?

It doesn't seem so. By merely postulating a new realm of entities, potential contradictions, one does not immediately aggregate explanatory power to the theory, neither do we gain in elegance, coherence, mathematical simplicity, or, what is more relevant, understanding. Our overall understanding of the situation is not improved by assuming a real contradiction dealt with by a paraconsistent logic. We are still left with questions about the collapse of the wave function when a measurement is made, for instance, and with questions as to why only one of them collapses, not to mention questions of how such a contradiction is to be understood (see also Arenhart and Krause [3]).

For another example, consider the case of nuclear models, as discussed by Morrison [19]. It is instructive to check how she summarizes the problem, in a few lines [19, p.351]:

> Here we have exactly the same phenomenon (the atomic nucleus and its constituents) modeled in entirely different ways depending on the data that needs explanation. However, in most cases these models go beyond mere data fitting to provide some type of dynamical account of how and why the phenomena are produced. But, because there are over thirty fundamentally different models that incorporate different and contradictory assumptions about structure and dynamics there is no way to determine which of the models can even

> be said to approximate the true nature of the nucleus. [...] Each of the models is both predictively, and in some cases, explanatorily successful in its particular domain but there is no way to build on and extend the models in a cumulative way. So, although we are able to extract information about the nuclear phenomena from these models we have no way of assessing the epistemic status of that information over and above its success in predicting certain types of nuclear phenomena.

That is, the trouble seems to be that the success of each kind of model is directly connected, some times, to the doubtful theoretical commitments of the models, commitments that other models directly contradict. It is not always easy to determine whether there is legitimate modeling or merely data fitting. As a result of the major difficulty in establishing a unified coherent picture, paraconsistency could be called as a helping hand. However, as Morrison [19, p.351] goes on to argue, this is clearly not a case where any kind of paraconsistency could help to save the day:

> What is perhaps significant for philosophical purposes is that this is not a situation that is resolvable using strategies like partial structures, paraconsistent logic or perspectivism. No amount of philosophical wizardry can solve what is essentially a scientific problem of consistency and coherence in theoretical knowledge.

The issue seems to be a general one. By changing logic to a paraconsistent one we by no means gain a better understanding of the underlying phenomena. That is, by claiming that something is legitimately both a wave and a particle does not add anything new to what we already knew about the theory: that some things are behaving in a way that is not completely coherently described by our current theory. Why is that so? Assuming contradictions adds nothing to answer that question. The claim could be more vividly illustrated by the case of moral dilemmas or in contexts of inconsistent laws, where practical consequences are even more immediate. By employing paraconsistent deontic logic we do not solve the legal problems we are involved in, we just grant that no explosion obtains from the contradictions. However, there is still not an issue about how to act morally in such contexts, or which law should be applied. Further: when witnesses of a crime contradict each other, we do not apply paraconsistent logic to tame the contradiction; that would be of no help! Paraconsistency leaves things where they are, and the same seems to be the case in science (anyway, let us mention that there are alternatives to deal with inconsistent contexts without the need of paraconsistent logics; see Michael [17]). The problem with contradictions in physics are not touched upon when we change logic; that is an issue that must be solved by the kind of theoretical advance that is provided by physics and theory development, something that is typically done by eliminating the contradictions, as Morrison has suggested.

So, it seems that by merely embracing contradictions we do not grant a better science, not even the maintenance of traditional scientific canons. The same could be said about the paradoxes of motion and change: do we gather a better understanding of those concepts by adhering to an inconsistent view of them, in place of the traditional one provided by the calculus? This does not seem to be the case. Contradictions leave us just where we are; *attempts to remove them provide for better understanding*, or at least it has been so for science (again, see Mosterín [20]).

4 Final Remarks: Where Can We Find True Contradictions?

So, in the end, the friend of true contradictions about the world will have heavy prices to pay. Assuming true contradiction in science forces us to adopt a whole bunch of uncomfortable positions. To keep the contradictions, we shall have to dictate the nature of the entities science deals with from the outside of science, from a philosophical position, ignoring the development of science as it eliminates contradictions. Furthermore, keeping with true contradictions we are obliged to adopt an inconvenient conservative and pessimistic attitude towards change in science. We may run the risk of imposing methodological obstacles to the development of science in order to keep our favorite aporias in the theory. Also, as we have seen, the main examples of possible aporias are not representative of the typical elimination of contradictions in science.

Furthermore, there is a rather negative ring on that argument; it places its bets on our inability to solve problems. Science, as a human enterprize, is the activity of searching for solutions for problems like those mentioned as examples of aporias. Keeping with contradictions would be like siding with the pessimist. This is clearly put by Priest [21, p.424]; even though he does not use da Costa's terminology, he makes it clear that when faced with an aporia, "we may find that there are no better ways to go. In which case, we may just have to conclude that the improbable [i.e., a true contradiction] is the case".

But more than recalling those problems, and linking with the previous quote by Priest, we would like to finish by remarking that the argument presented by da Costa favoring true contradictions about the world has striking similarities with another style of argument by Priest [22] in favor of adopting true contradictions. Let us consider the most typical case first, the paradoxes of self-reference such as the Liar. Concerning them, Priest says:

> Here we have a set of arguments that appear to be sound, and yet which end in contradiction. *Prima facie*, then, they establish that some contradictions are true. Some of these arguments are two thousand and a half years old. Yet, despite intensive attempts to say what is wrong with them in a number of logical epochs, including our own, there are no adequate solutions. (Priest [22, p.83])

He then goes on to compare the self-reference paradoxes with Zeno's paradoxes. According to Priest [22, p.83], there is agreement about what the solution to Zeno's paradoxes are, with only minor details still left for dispute. In the case of paradoxes of self-reference, however, no such consensus on a solution is forthcoming, and this suggests that "trying to solve them is simply barking up the wrong tree: we should just accept them at face value, as showing that certain contradictions are true" (Priest [22, p.83]).

Further arguments on the same line are provided by Priest on what concerns moral dilemmas and rational dilemmas (see [23, chap.13] and [22, chap.6]). Those dilemmas are not explicit contradictions (just as being a wave and a particle is not), but they do sometimes imply contradictions (just as being a wave and a particle do, under reasonable assumptions). Their existence, as Priest says, "is simply a fact of life" ([22, p.111]). It should come as no surprise now that the solution to those problems, according to Priest, is to live with the dilemmas, just as we should do with the contradictions involved in the semantic paradoxes. That is: in the absence of a non-controversial solution, embrace the contradiction. Of course, the practical question is left untouched, but, as dialetheists see things, contradictions are here to stay even in the presence of rational dilemmas:

> If there are such things, the next obvious question is what one should do if one finds oneself in one. What one should do, is, of course, the impossible. But one can't do that. Rationally, one way or other, one is damned. *C'est la vie*. (Priest [22, p.115])

Notice that except for the fact that Priest and da Costa disagree on whether Zeno's paradoxes exemplify aporias, they agree that the Liar paradox (and paradoxes of self-reference in general) is an aporia and that the fact that attempts at solving it have taken so much controversies for so much time means that the contradiction may be here to stay! Probably the same would be true on what concerns moral and rational dilemmas. So, the arguments are completely analogous, with the proviso that Priest does not advance his version of the argument as including contradictions in concrete reality (although the Liar is said to be true in this world, the actual world). The upshot is that dialetheists have a common strategy to motivate their position, and it always appeal to our failure in reaching consensus about consistent solutions to certain paradoxes. What da Costa did was to advance the argument in complete generality, aiming also at the world itself with cooperation of science, while Priest restricts the argument, it seems, to the typical cases involving language, self-reference, and semantic concepts.

Now that we have seen that adopting contradictions as true in the real world is not warranted, perhaps a similar case could be made against the same argument as advanced by Priest (that is, the same argument now restricted to paradoxes involving semantic and semiotic notions, as well as moral and rational ones). By treating the arguments with distinct application spheres we may learn more about the scope of dialetheism and, hopefully, reach a consensus that dialetheism itself is not required. Contradictions may be just dismissed as false in the end, even though they should be recognized as one of the main forces driving progress in science. Interesting as they are, those lines of investigation are an issue for another work.

Acknowledgements We would like to thank two anonymous referees for their constructive criticisms and comments which helped to improve the paper.

References

1. Ardourel, V. 2015. A discrete solution for the paradox of Achilles and the tortoise. *Synthese* 192 (9): 2843–2861.
2. Arenhart, J.R.B. 2012. Ontological frameworks for scientific theories. *Foundations of Science* 17: 339–356.

3. Arenhart, J.R.B., and D. Krause. 2016. Contradiction, quantum mechanics, and the square of opposition. *Logique et Analyse* 59 (235): 301–315.
4. Barker, S., and M. Jago. 2012. Being positive about negative facts. *Philosophy and Phenomenological research* 85: 117–138.
5. Beall, J. C. 2004. Introduction: At the intersection of truth and falsity. In *The Law of Non-Contradiction: New Philosophical Essays*, ed. G. Priest, J.C. Beall, and B. Armour-Garb, 1–19. Oxford: Clarendon Press.
6. Beall, J.C. 2009. *Spandrels of truth*. Oxford: Oxford University Press.
7. Berto, F. 2012. How to rule out things with words: Strong paraconsistency and the algebra of exclusion. In *New waves in philosophical logic*, ed. G. Restall, G. K. Russell, 169–189. New York: Palgrave Macmillan.
8. Bobenrieth, A. 2007. In *Paraconsistency and the Consistency or Inconsistency of the World*, ed. J.Y. Béziau, W. Carnielli, and D. Gabbay, Handbook of Paraconsistent Logic, 493–512. London: College Publications.
9. da Costa, N.C.A. 1997. *Logiques Classiques et Non Classiques. Essai sur les fondements de la logique*. Paris: Masson.
10. da Costa, N.C.A. 2002. *Logic and Ontology. Principia* 6 (2): 279–298.
11. da Costa, N.C.A., and C. de Ronde. 2013. The paraconsistent logic of superpositions. *Foundations of Physics* 43 (7): 845–858.
12. Deguchi, Y., J.L. Garfield, and G. Priest. 2013. The contradictions are true – and it's not out of this world! a response to Takashi Yagisawa. *Philosophy East and West* 63 (3): 370–372.
13. de Ronde, C. 2015. Modality, potentiality, and contradiction in quantum mechanics. In *New Directions in Paraconsistent Logic*, ed. J.-Y. Béziau, M. Chakraborty, and S. Dutta, 249–265. New Delhi: Springer.
14. Grim, P. 1991. *The Incomplete Universe: Totality, Knowledge, and Truth*. Cambridge: MIT Press.
15. Horsten, L. 2015. One hundred years of semantic paradox. *Journal of Philosophical Logic* 44: 681–695.
16. Mares, E. D. 2004. Semantic dialetheism. In *The Law of Non-Contradiction: New Philosophical Essays*, ed. G. Priest, J.C. Beall, and B. Armour-Garb, 264–275. Oxford: Clarendon Press.
17. Michael, M. 2016. On a most telling argument for paraconsistent logic. *Synthese* 193 (10): 3347–3362.
18. Morganti, M., and T.E. Tahko. 2017. Moderately naturalistic metaphysics. *Synthese* 194 (7): 2557–2580.
19. Morrison, M. 2011. One phenomenon, many models: inconsistency and complementarity. *Studies in History and Philosophy of Science* 42: 342–351.
20. Mosterín, J. 2011. The role of consistency in empirical science. *Manuscrito: revista internacional de filosofia* 34(1):293–305.
21. Priest, G. 1998. What is so bad about contradictions? *The Journal of Philosophy* 95 (8): 410–426.
22. Priest, G. 2006. *In Contradiction: A Study of the Transconsistent*, 2nd ed. Oxford: Oxford University Press.
23. Priest, G. 2006. *Doubt Truth to be a Liar*. Oxford: Oxford University Press.
24. Priest, G. 2014. Contradictory concepts. In *Logic, Reasoning, and Rationality*, ed. E. Weber, D. Wouters, and J. Meheus, 197–215. Berlin: Springer.
25. Priest, G., and R. Routley. 1989. The philosophical significance and inevitability of paraconsistency. In *Paraconsistent Logic: Essays on the Inconsistent*, ed. G. Priest, R. Routley, and J. Norman, 483–537. Munich: Philosophia Verlag.
26. Tahko, T. 2009. The law of non-contradiction as a metaphysical principle. *Australasian Journal of Logic* 7: 32–47.

The Possibility and Fruitfulness of a Debate on the Principle of Non-contradiction

Luis Estrada-González and Maria del Rosario Martínez-Ordaz

Abstract Five major stances on the problems of the possibility and fruitfulness of a debate on the principle of non-contradiction (PNC) are described: Detractors, Fierce supporters, Demonstrators, Methodologists and Calm supporters. We show what Calm supporters have to say on the other parties wondering about the possibility and fruitfulness of a debate on PNC. The main claim is that one can find all the elements of Calm supporters already in Aristotle's works. In addition, we argue that the Aristotelian refutative strategy, originally used for dealing with detractors of PNC in *Metaphysics*, has wider implications for the possibility and fruitfulness of an up-to-date debate on PNC, at least in exhibiting some serious difficulties for the other parties.

1 Introduction

Up to now, five major stances on the two long-standing interconnected problems of the possibility and fruitfulness of a debate on the Principle of Non-contradiction (PNC henceforth) can be recognized, namely:

- *Detractors* are ready to give PNC up; its relatively straightforward failure would be enough ground to the possibility of disputing it. Aristotle in *Metaphysics* [1] construed several ancient thinkers as detractors of PNC, notably Heracliteans, and today the most visible effort is dialetheism as in [27].

The original version of the chapter was revised: The name of the author has been corrected from "First Name: María and Family Name: del Rosario Martínez-Ordaz" to "First Name: Maria del Rosario and Family Name: Martínez-Ordaz". The correction to the chapter is available at https://doi.org/10.1007/978-3-319-98797-2_15

L. Estrada-González · M. R. Martínez-Ordaz (✉)
Institute for Philosophical Research, UNAM, Mexico City, Mexico
e-mail: martinezordazm@gmail.com

L. Estrada-González
e-mail: loisayaxsegrob@gmail.com

W. Carnielli and J. Malinowski (eds.), *Contradictions, from Consistency to Inconsistency*, Trends in Logic 47, https://doi.org/10.1007/978-3-319-98797-2_3

- *Fierce supporters* say that it holds universally and that it is so certain and basic that it cannot be shown to hold on any more basic grounds. This is the line famously held by Lewis ([21, 22]).
- *Demonstrators* such as Kant ([18]), Boole ([7]), Russell ([30]) or Priest ([26]) are more open to debate, not necessarily because they are ready to give PNC up, but because they think it can be demonstrated in terms at least as secure as the PNC itself.
- Others, *Methodologists*, say that PNC can not only be discussed, but accepted or rejected just as any other claim, namely using methodological principles of rational choice. This view is espoused for example by Bueno and Colyvan ([8]) and Priest ([26]).[1]
- *Calm supporters* say that PNC has usually been formulated in some strong ways and that Detractors are rightly attacking those formulations yet they should accept a very basic form of PNC as to ensure the intelligibility of their proposals and criticisms. This is basically the proposal recently outlined by Berto ([4, 5]) and practiced e.g. by Tahko ([32]).

However, in the emergence of the last stance, the focus has been on how they interact with Detractors. In this paper we show what Calm supporters have to say on the other parties wondering about the possibility and fruitfulness of a debate on PNC. The main claim is that one can find all the elements of calm supporters already in Aristotle's works, and that his way of dealing with detractors of PNC in *Metaphysics* has wider implications for the possibility and fruitfulness of an up-to-date debate on PNC. Aristotle's way to refute Detractors not only would do that, but also shows how to conduct a debate about PNC even if it is certain and holds universally, against Fierce supporters; why it is not demonstrable, against Demonstrators, and not even subject to settlement or rejection through methodological principles, against Methodologists.[2]

Two important remarks are in order here. First, we will not attempt a defense of Calm supporters; we merely want to show how the Aristotelian refutative strategy can be used beyond its original target, the ancient Detractors, and that it succeeds at least in exhibiting some serious difficulties for the other parties. A more thorough examination of each of them is left for further work. Secondly, this paper does not constitute a merely historical reconstruction of the Aristotelian stand point, but it is an attempt to provide a better understanding of the current debate on PNC. Thus, some historical and exegetical details about Aristotle's work will be left aside in order to emphasize its import for more contemporary views regarding PNC.

---

[1] Advocators of "anti-exceptionalism about logic" —for example, Williamson ([37, 38]) and Hjortland ([23])— can be regarded as methodologists, and their views surely would have implications for the debate here discussed. Note also that Priest appears as a detractor and a demonstrator, a contradiction that, for him, would show that the debate would be better approached through the methodological principles of rational choice. This would be more extensively discussed in Sects. 6 and 7 below.

[2] The only caveat was that Aristotle aimed at establishing a version of PNC stronger than his refutations of detractors allowed to conclude but, as we have mentioned, this is not a problem for Calm supporters in general.

The plan of the paper is as follows. In Sects. 2–4 we introduce some basic Aristotelian terminology which serves as background for what follows. Although at some points we do not steer clear from exegetical discussion, our main interest lies on suitable logical reconstructions of Aristotle's views, useful for the overall issue about the debate on PNC than in an exposition completely sound to the ears of a scholar on them. In Sect. 2 we distinguish several kinds of principles of non-contradiction present in *Metaphysics Γ* and some of its properties. Section 3 is devoted to make more precise about the semantic version of PNC which will be discussed throughout the paper. In Sect. 4 we discuss Aristotle's notions of demonstration and refutation, and show how they help to deal with the anti-debate stance of Fierce supporters like Lewis. In Sect. 5 we reconstruct one of Aristotle's refutations of Heracliteans. This will prove useful for showing what is wrong with the approaches of both Demonstrators —issue dealt with in Sect. 6— and Methodologists —analyzed in Sect. 7—.

2 Aristotle's Principles of Non-contradiction

Aristotle discussed at least three closely interrelated versions of the Principle of Non-contradiction (PNC):

An *ontological* version (oPNC):	It is impossible for a thing to be and not to be (*Metaphysics* [1] 996b30; 1005b18-21)[3];
a *semantic* version (sPNC):	Given a proposition and its negation, they cannot be both true (cf. *Metaphysics* [1] 1011b11-14);
a *doxastic* version (dPNC):	It is impossible to believe contradictions (pairs consisting of a proposition and its negation; *Metaphysics* [1] 1005b23f).[4]

Aristotle claimed —although not always with this precise jargon, of course— that PNC is fundamental or a first principle because of its

Groundingness	things could not be if it did not hold (cf. *Metaphysics* [1] 1006a28-1007b18; 1007b18-1008a2; *Physics*, ch. 3);
Universality	it holds in all situations (cf. *Metaphysics* 1007b18-1008a2);
Certainty	it is the most certain and best known principle (cf. *Metaphysics* 1006a5-1006a18);
Indemonstrability	it is presupposed in every demonstration (so it is itself indemonstrable) (cf. *Metaphysics* 1006a5-

[3] An ancestor of this is found in Plato, *Republic* [24] 436b: "It is obvious that the same thing will never do or suffer opposites in the same respect in relation to the same thing and at the same time."

[4] Also found in Plato. See *Republic* [24] 602e8-9: "And haven't we said that it is impossible for the same [person] to think/imagine at once/simultaneously opposite [things] about the same [things]?"

1006a18; 1062a1ff; 1062a30);
[5] should it be violated, communication, and in general, purposeful action, would be impossible (cf. *Metaphysics* 1008a8-1006b31).

Although all of the features sound well for oPNC, it is not the same for the other versions. For example, it is not clear that *Groundingness* could be a feature of sPNC or dPNC. Also, given the primacy of metaphysical issues for Aristotle, it is not clear whether sPNC could be demonstrated through, or somehow based on, the oPNC.

In this paper we are interested in capturing in a systematic the state of the art of the debate regarding PNC; for that reason, from now on, we will be focused mostly in a version of sPNC since it has recently been emphasized that the issue of the possibility of a debate can be more easily treated by discussing *Indemonstrability* ([4, 5, 26, 27, 32]). So, all of the features presented above should be taken as features of a semantic first principle. For example, such a first principle would ground all semantics, would hold for all semantic expressions, it would be more certain than any other semantic principle and it could not be demonstrated on the basis of other semantic principles. Thus, in what follows 'PNC' will stand for a version of sPNC that will be clarified in the following sections.

3 Some Distinctions

It is common to muddle (i) principles concerning the structure of truth values, (ii) principles concerning the relations between propositions and truth values and (iii) the theoremhood or otherwise of certain propositions. For instance, *Bivalence* is a principle concerning the structure of truth values: There are exactly two of them.[6] *Falsity of at least one contradictory* is a principle concerning the relations between propositions and truth values: One of a proposition and a negation of it is false. $p \vee \neg p$ is a proposition that may or not be a logical truth. These three things are independent of each other.

The situation is analogous in the case of PNC:

No values-gluts	No truth value can be described as true-and-false;
(Interpretational)	(a) only one of a proposition and a negation of it is true, or
Non-contradiction:	(b) no proposition is both true and false;
Logical falsehood of	$A \wedge \neg A$, for any A
Theoremhood of	$\neg(A \wedge \neg A)$, for any A.

[5] We do not know a simple expression to name this property, not even in our mother-tongue, and certainly we do not dare piss the owners of English off by inventing one, hence the blank space.

[6] And it must not be confused with *Non-referential gap*: There is no truth value describable as neither-true-nor-false. These are independent principles. *Bivalence* might not hold because, say, there is a third value, but it is not "neither-true-nor-false" but rather, say, "true-and-false".

Again, these four claims are independent of each other.[7] Thus, neither the mere failure in a logic of $A \wedge \neg A$ of being always false, nor both the failure of $A \wedge \neg A$ being always false and the theoremhood of $\neg(A \wedge \neg A)$, as happens in numerous contemporary logics, is enough evidence that Aristotle's PNC has been effectively violated. In addition, the fact that there can be a deduction of $\neg(A \wedge \neg A)$ in a logic does not suffice as evidence for its demonstrability.

In most of *Metaphysics* Γ Aristotle discusses whether the principles on the structure of truth values or concerning the relations between propositions, which together entail the wished properties of formulas. However, even *No values-gluts* and (Interpretational) *Non-contradiction* would be, with certain additional assumptions, instances of a more basic semantic principle, stated at the beginning of his discussion of Detractors (cf. *Metaphysics* [1] Γ, 1006b11):

Semantic exclusion: An expression cannot have two mutually exclusive semantic properties.[8]

Then, in what follows 'PNC' will stand for *Semantic exclusion* and only derivatively for (Interpretational) *Non-contradiction* or *Non-referential glut*. We would not bother with distinguishing when *Semantic exclusion* is instantiated for propositions or for descriptions of truth values; context would be enough guide, and in any case *Semantic exclusion* would be doing most of the work.[9] Its proper scope will be clear in Sect. 5.

4 Aristotle on Demonstration, Refutation and the Possibility of a Fruitful Debate on PNC

According to Aristotle, "a deduction (*sullogismos*) is [a] speech in which, certain things having been supposed, something different from those supposed results of necessity because of their being so." (*Prior Analytics* [3] 24b18-20) Meanwhile, a *demonstration* is for him a deduction in which the things supposed are known to be true, primary, immediate, better known and prior than the resulting thing and then the resulting thing is known (cf. *Posterior Analytics* 71b10-25).[10] "Certainty", "immediacy", "priority" or "better" knowledge does not mean, for Aristotle, self-

[7] A key ingredient in classical logic is *Functionality*: (a) Every proposition has one value (b) but only one. In a language with the usual ingredients (conjunction, disjunction, etc. as certain functions) this, together with *Bivalence*, *Falsity of at least one contradictory*, *No values-gluts* and either version of (Interpretational) *Non-contradiction*, suffices to give all the usual valuations of classical logic.

[8] For more details on this principle and its importance on Aristotle's overall argument against Detractors of PNC in *Metaphysics* Γ, see [35], where the principle is taken as saying that the meaning of a term is unique, definite and determinate. Nonetheless, we take the mutual exclusiveness of certain semantic properties as a more appropriate reconstruction of Aristotle's claims on the uniqueness, definiteness and determinateness of meanings and their role in his refutations.

[9] For a guided tour on even more versions of PNC, see [16].

[10] Whereas it can be debated whether Aristotle's notion of deduction allows reflexivity (A is deducible from A) because of the "something different from [the things] supposed results" clause,

evident knowledge. He distinguishes between what is better known for someone from what is better known "in itself" or "in nature", and says that he means the latter in his characterization of demonstration. Thus, that someone does not find PNC true at first sight does not count against, say, its *Certainty* (cf. *Posterior Analytics* 71b34-72a4; 76a18-22).

Aristotle holds that PNC, in either of his versions, is a "principle of demonstration" and, thus, there is no demonstration in which PNC and only PNC is the conclusion; it is indemonstrable (*Metaphysics* 1006a5-1006a18). There is an important remark by Aristotle as to in what sense PNC is a principle of demonstration:

> No demonstration assumes that it is impossible to assert and deny at the same time —unless the conclusion too is to be proved in this form. Then it is proved by assuming that it is true to say the first term of the middle term and not true to deny it. It makes no difference if you assume that the middle term is and is not (…). (*Posterior Analytics* 77a10-14)

Aristotle's point can be illustrated in this way. If all As are Bs and all Bs are Cs, it follows that all As are Cs (this is just a *Barbara* form of argument), even if some As are also not Bs. After all, they are still Bs as well, and so Cs. But if one wants to conclude that all As are Cs and only that, i.e. not also that not all As are Cs, then PNC has to be assumed.

Let us generalize this point given that the occurrence of negation here would be rather an instance of the exclusion mentioned in the previous section: Whenever one wants to demonstrate something but not also something excluded by it, one has to presuppose PNC. If one presupposes it in no demonstration, one always could demonstrate something but also something excluded by it, hence all exclusions and probably everything. This remark will prove crucial to assess the attempts to demonstrate PNC.

Given its *Universality*, PNC cannot be refuted, and given its *Indemonstrability* it is, well, indemonstrable, so seemingly there would be no place for debate about it. When responding to an invitation to contribute a piece to an anthology on the debate about PNC, Lewis [22, 176] explicitly rejected the possibility of such a debate:

> I'm sorry; I decline to contribute to your proposed book about the 'debate' over the law of non-contradiction. My feeling is that since this debate instantly reaches deadlock, there's really nothing much to say about it. To conduct a debate, one needs common ground; principles in dispute cannot of course fairly be used as common ground; and in this case, the principles *not* in dispute are so very much less certain than non-contradiction itself that it matters little whether or not a successful defence of non-contradiction could be based on them.[11]

demonstration certainly does not allow it: A would have to be both more certain (for it is a premise in the demonstration) and less certain (for it is demonstrated) than itself.

[11] This was a permanent stance:

> No truth does have, and no truth could have, a true negation. Nothing is, and nothing could be, literally both true and false. This we know for certain, and a priori and without any exception for especially perplexing subject matters. (…) That may seem dogmatic. And it is: I am affirming the very thesis that [the foes of the PNC] have called into question and —contrary to the rules of debate— I decline to defend it. Further, I concede that it is indefensible against their challenge. They have called so much into question that I have no

Lewis' argument is intended neither as a demonstration nor as a justification of PNC, but rather as a plea for its adoption as to ensure the possibility of any debate.[12] He even says in another place: "I think this [the impossibility of a debate on PNC] calls in question the very idea that philosophy always can and should proceed by debate —itself a heretical view, likely to be vigorously opposed." (Quoted in [27, xix].)

Maybe Lewis is right in that not everything can be debated, but it is not clear that no debate on PNC is possible. One problem with Lewis' argument is that it gives much weight to the *Certainty* of PNC. Lear [20, Ch. 6.4] has correctly highlighted that, for Aristotle, the *Certainty* of PNC does not make it self-evident, so it can be honestly challenged, because Detractors can fail to acknowledge some of its other properties. The story does not end, though, with the supporters of PNC acknowledging, say, the *Universality* and *Certainty* of PNC and with Detractors failing to grasp them, as Lewis' argument seems to suggest.

Yet another problem is that Lewis seems to restrict the field of debate about PNC to just either demonstrating it ("defend it") or refuting it. Since PNC holds, refuting it is not an option. And since it cannot be demonstrated, because there is nothing more basic from which it could be demonstrated, no debate would be possible. But Aristotle recognized another way to debate, namely defending PNC through the *refutation* of their opponents (*Metaphysics* 1006a11ff; 1062a2). A refutation of a claim (possibly endorsed by someone) is for him, typically, the deduction of a contradiction from that claim (*Prior Analytics* [3] B20, 66b11). However, in the context of debating PNC, deducing a contradiction from the claims of those who do not endorse it would not work, on pain of blatantly begging the question. Rather, a refutation in this context would exhibit something incompatible with the premises of the Detractors, or something patently absurd by their own lights, so it "silences" them (*Metaphysics* [1] 1009a17-20).[13] Aristotle would try to show that the Detractors' most important

> foothold on undisputed ground. So much the worse for the demand that philosophers always must be ready to defend their theses under the rules of debate. [21, 101]

And in a letter interceding for the publication of Priest's *In Contradiction*:

> Many people will think that it is an easy thing to refute Priest's position, decisively and in accordance with customary rules of debate. It is not an easy thing. I myself think that it is an impossible thing: so much is called into question that debate will bog down into question-begging and deadlock. (Quoted in [27, xix].)

In what follows we will use interchangeably 'x is debated', 'x is disputed', 'the possibility that x does not hold is entertained' and even 'x is defended' in this reconstruction.

[12] Field [14] and Boghossian [6] call a principle "default reasonable" if it is regarded as being good at leading to true beliefs and avoiding error (hence reasonable) and it is done so without first adducing evidence or argument in its favor (hence default). We do not think PNC would be default reasonable in this sense for Lewis: It is not merely good at leading to true beliefs and avoiding errors, but is the sine qua non to draw a distinction between true beliefs and errors at all; also, it is not merely default, but already the distinction between what counts as evidence and an argument in favor of something presupposes it.

[13] See [9, Ch. 5] for further details.

premise is that whatever they say, even the smallest of words, is regarded by them as intelligible, and so their denial of PNC should imply the ultimate unintelligibility of what a detractor says. Said otherwise, a refutation of Detractors takes the form of an exhibition that the intelligibility of what they say assumes the validity of (at least a form of) PNC, as in 'Some of my most basic expressions mean this and only this, nothing else'.

Thus, debate can proceed according to Aristotle since PNC and its features might not be obvious; and although it cannot be demonstrated nor proven false, it can be defended through refutation. That the validity of PNC is essential to the meaningfulness of certain expressions can be made patent to Detractors, and with it, others of PNC's features, like its *Certainty* and *Universality*, or any other of its features that might be in dispute, can also be made patent to those who call them into question.[14]

Thus, what one can learn from Aristotle is that a debate is *possible* if each party has at least entertained the possibility that the other's position might have some intelligibility. Lewis is clearly demanding something stronger: that they all agree on some substantial principles. On the other hand, a debate would be considered as *fruitful* if it satisfies any of the following:

(Weak fruitfulness) One of the parties is moved to actually find the other's position intelligible (nearly) as a whole in virtue of the exchange.

(Strong fruitfulness) One of the parties is moved to modify their position in virtue of the exchange with the other one.

Again, Lewis demands at least the stronger version of fruitfulness for a debate on the PNC.

5 Aristotle's Refutation of Heracliteans

So, debate with Detractors is possible through refutation. The result will always be establishing (some feature of) PNC, there is no much to debate about that, by exhibiting the detractor's ultimate commitment to PNC. One of the refutations of Heracliteans, found in *Metaphysics* 1008a34-1008a38, shows particularly well the Aristotelian strategy to make Detractors aware of their commitment to some form or feature of PNC. But it shows not only that debate is possible, against Fierce supporters, and that Detractors are wrong, at least in this specific case, but this refutation has all the ingredients that show why Demonstrators and Methodologists would be on the wrong track on the issue of debating PNC and we will deploy them in the next sections.

Before reconstructing the refutation of Heracliteans, a note on terminology is in order. For Aristotle, logic and logical notions have strong multi-agent dialogical components. For example, we have seen that deduction is a kind of speech, and so are demonstration and refutation. However, we have assumed that it is exegetically

[14] Thus, *pace* Priest [26], we take sides with those who say that some of Aristotle's refutations establish not necessarily PNC itself or at least its *Universality*, but something else about it, even if that was not Aristotle's aim. See [34] for further discussion.

safe for our present concern to rely on what Dutilh Novaes [13, 123] calls "the internalization of the interlocutors" in the modern proof- or model-theoretic analyses of logic. For example, a typical reductio in modern format exhibits a form of refutation where the interlocutors are somehow internalized in the formal machinery. So, we reconstruct cognitive-dialogical expressions like

(CD) The (group of) agent(s) s accepts/rejects/asserts/denies/believes/etc. what the (group of) agent(s) s^* accepts/rejects/asserts/denies/believes/etc.

as

(RCD) The (group of) agent(s) s accepts/rejects/asserts/denies/believes/etc. $E(A)$.

where '$E(A)$' is a reconstruction of what s^* accepts/rejects/asserts/denies/believes/etc. in terms of evaluations of the proposition A accepted/rejected/asserted/denied/believed/etc. by s^*. Even more frequently, we reconstruct (CD) internalizing all interlocutors as

(ICD) $E(A)$

i.e., "The evaluation of A is such and such" as a step in an s-ian argument about s^*'s position. We will rely on such internalization later when talking about speech acts like assertion and denial or even cognitive states like acceptance and rejection.[15]

Aristotle's refutation reads as follows: "Again, if when the assertion is true, the negation is false, and when this is true, the affirmation is false, it will not be possible to assert and deny the same thing truly at the same time. But perhaps they might say this was the very question at issue." (*Metaphysics* 1008a34-1008a38) We provide here a more or less thorough reconstruction of the argument to make as clear as possible its basic premises[16]:

Basic premises:

(A1) For every valuation v and proposition A, if $v(A) = \top$ then $v(\neg A) = \bot$.
(A2) For every valuation v and proposition A, if $v(\neg A) = \top$ then $v(A) = \bot$.
(A3) For every valuation v and propositions A and B, if either $v(A) \neq \top$ or $v(B) \neq \top$ then $v(A \wedge B) \neq \top$.
(A4) For every valuation v and proposition A, either $v(A) = \top$ or $v(A) \neq \top$.
(A5) There are exactly two truth values, $\top \neq \bot$.

Rest of the argument:

(A6) Suppose $v(A) = \top$
(A7) Then $v(\neg A) = \bot$ (from (A1))
(A8) Then $v(\neg A) \neq \top$ (from (A7) and (A5))

---

[15] Priest [26, 14f] does not regard as refutations some arguments where apparently there is no other agent saying anything on the basis that Aristotle required Detractors to say something. This internalization would help to explain the fact that there could be a refutation even if there were nobody saying anything: Detractors' saying has been internalized in the supporter's argument.

[16] We would like to emphasize that the following reconstruction of Aristotle's refutation strategy is not an attempt to make an exegetically perfect reconstruction —considering, for example, the fact that it is described in terms of a contemporary semantics. The main purpose of this reconstruction is to show in a comprehensive manner how the strategy works. We leave to the Aristotelian scholars to decide whether it matches perfectly the original proposal.

(A9) Then $v(A \wedge \neg A) \neq \top$ (from (A8) and (A3))
(A10) Now, suppose $v(A) = \bot$
(A11) Then $v(A) \neq \top$ (from (A10) and (A5))
(A13) Then $v(A \wedge \neg A) \neq \top$ (from (A11) and (A3))
(A13) Make analogous proofs about the truth values of $\neg A$ until you have covered all cases.
(A14) Then, for every valuation v, $v(A \wedge \neg A) \neq \top$ (from (A1) to (A13))

Aristotle says: "But perhaps they might say this was the very question at issue." What is the "very question at issue", what do 'they' and 'this' stand for? Dancy [12] speaks of a single (fictional) target of Aristotle's refutations and calls him "Antiphasis". However, Aristotle's refutations are directed at different possible foes of PNC that is better to keep separated for the purposes of analysis: Dialetheists, for whom some but not all contradictions are true; Heracliteans, for whom all contradictions but not other propositions, especially their conjuncts, are true; trivialists, for whom everything is true; and even logical nihilists, for whom nothing is true.[17] Priest [26, 38] seems to think that the argument studied in this section is directed against a supporter of dialetheism or trivialism, and that it is unsuccessful against them. However, the thesis under discussion is "All contradictions (and only them) are true", which comes since *Metaphysics* 1007b26 and is attributed to the Heracliteans (cf. *Metaphysics* [1] 1005b25, 1012a25), and which we will call "logical Heracliteanism".[18] The argument does not bear directly upon dialetheism, because it does not imply logical Heracliteanism. And although trivialism implies logical Heracliteanism, it is not the explicit target of the argument.[19]

Logical Heracliteanism is based on two theses. According to *metaphysical Heracliteanism*, everything is in state of flux at every moment; so, and this is *semantic Heracliteanism*, a thing cannot be described truly to be an F because it would be to fix it, and the same considerations are made for not being an F, but it can be described truly and fully as both being an F and not being an F. Hence all contradictions, but none of their components, are true, because it would be the only way to capture the changing nature of things.[20]

[17] And at least in one case, Aristotle distinguishes between Heracliteanism and dialetheism because they need to be refuted in different ways, cf. *Metaphysics* 1008a7-12.

[18] Whether these views should actually be attributed to the historical Heraclitus is a moot point; see [15, Ch. 5]. However, the interest here lies more in the position itself than in the correctness of Aristotle's scholarship.

[19] The problem is that Priest sometimes conflates trivialism —"Everything is true"— with a version of Heracliteanism —'All contradictions are true'— (cf. [28, 131], although sometimes he says that he is aware that the identification depends on certain assumptions, notoriously 'and'-elimination (see [26, 56]. Aristotle too thought that semantic Heracliteanism could be equated with trivialism, but he was cautious: "The doctrine of Heraclitus, that all things are and are not, seems to make everything true (…)" (*Metaphysics*1012a25).

[20] In proof-theoretic terms, Heracliteans would not accept conjunction elimination when the premise is a contradiction, and they would accept conjunction introduction when the conjuncts are not true but only if they are contradictories. In model-theoretic terms, it may be that $v(A) \neq \top$ or $v(B) \neq \top$ even if $v(A \wedge B) = \top$. So, when the conjuncts are contradictories, semantic Heracliteanism's conjunction

So "the very question at issue" is what allows going from the truth conditions of negation to the conclusion, i.e. the idea that a simultaneous affirmation of two affirmations cannot be true if one of them is not true, (A3). In order to avoid begging the question and complete the refutation, Aristotelians have to do another move, and Aristotle has provided the pattern for that in his previous refutations (1006a18-1088a33). Heracliteans say that (H) It is not the case that, for every propositions A and B, if either $v(A) \neq \top$ or $v(B) \neq \top$ then $v(A \wedge B) \neq \top$. But (H) has the form Not-(A3). Moreover, according to Heracliteans, (H) must fail to be true since only $(H) \wedge \neg(H)$ could be true. So, if Heracliteans want to say that their principle Not-(A3) is true, they have to say that it is true only together with (A3). And indeed it has to be so for any Aristotelian premise.[21]

However, this is not a great victory yet. Aristotelians would prefer Heracliteans to accept only their premises and conclusion as true, but they merely get that Heracliteans have to take both them and their negations as true. To say that Heracliteans cannot do that is to assume PNC. For the argument to run, Aristotelians need Heracliteans wanting to reject (A3) by saying that it is just false while only Not-(A3) is true. But according to Aristotle, that is exactly what Heracliteans want to do, which would undermine their own position, so the desired result for Aristotelians would obtain.[22] A shorter, less convoluted refutation of Heracliteanism that we have not found in Aristotle's works would be as follows: If all and only contradictions are true, then there is some non-contradiction that is true, namely the Heraclitean thesis itself ("All and only contradictions are true"). Therefore, not all contradictions and only them are true.

In any case, if Heracliteans wanted to say that only Not-(A3), their (H), is true, they would have to endorse some form of PNC, the *minimal semantic principle of non-contradiction* (MSPNC):

(MSPNC) "For some semantic property P, family of semantic properties $Q_1, \ldots, Q_n$ and some expression e of a non-trivial theory T, it is impossible that P holds good and that any of $Q_1, \ldots, Q_n$ holds good for e."

For simplicity, the disjunction of the Q_is will be collectively denoted by '*Not-P*'. This is like the principle of *Semantic exclusion* mentioned in Sect. 3, but with explicit quantifiers. The argumentation against Heracliteans makes clear the rationale

resembles the relevance logics' fusion connective, and thus is not so odd by contemporary lights. Besides, in certain sense, semantic Heracliteanism is dual to non-adjunctivism, which is a thesis found in some paraconsistent logics like the earliest one, Jaśkowski's: In non-adjunctivism, in general $v(A \wedge \neg A) \neq \top$ even if $v(A) = v(\neg A) = \top$. Non-adjunctivism was used in some of the earliest attempts to make sense of impossible worlds; see [29].

[21] An argument similar to this one is presented in *Metaphysics* K (see 1062b2ff).

[22] Again according to Aristotle, Cratylus held that even contradictions over-fix reality so they must be altogether false, like their conjuncts. However, he noticed the problem faced by the Heraclitean and did not make the mistake of trying to philosophize about that, so he attempted to threw the ladder out after stating his position and ceased to do virtually any philosophical statement (only "virtually" because he kept denying with his forefinger and that still causes problems). See *Metaphysics* 1010a10ff.

behind MSPNC. In order for a theory T to rule out something, some of its expressions must mean something and nothing else incompatible with that something, and when certain contradictions are true, some of the expressions involved in them must mean something else incompatible with what they supposedly mean. If logical Heracliteanism is expressible at all, it is false, because its expressibility requires either that some non-contradictions are true, like Not-(A3) alone, or that somehow only (A3) and Not-(A3) together mean Not-(A3), but the meaning of Not-(A3) does not seem to be meaning of a contradiction.[23]

MSPNC is decidedly weaker than some might expect. But in fact most of Aristotle's refutations do not succeed beyond establishing restricted versions of PNC, like MSPNC. It only holds for *some* expressions and *some* semantic properties, since form the fact that Heracliteans need to recognize at least one non-contradiction as true does not follow that only non-contradictions may be true, or more simply, the negation of "All contradictions and only them are true" is not "No contradiction is true". Nonetheless, the scope of MSPNC is not a negligible one. Holding for all theories which aim to rule something out from it just leaves one theory out, trivialism. So *Universality* might not hold in the sense that MSPNC does not apply to any expressions, but it holds in the sense that virtually every theory must have such incompatible semantic properties.[24]

6 The Consequences of Refutations for Demonstrators

Aristotle's argument in the previous section gives a clear example of why he does not consider his refutations as demonstrations. Without (MS)PNC, there is no obstacle for considering Not-A_n for any of the Aristotelian premises A_n, and then either proving Not-PNC, or both PNC and Not -PNC, but not only PNC. So (MS)PNC has to be presupposed in every demonstration that wants to demonstrate one thing and only it, not also something excluded by it.

Demonstrations of PNC like Kant's [18], Boole's [7, Ch. 3] or Russell's (cf. [36, *362 3.24])[25] fall short for the same reason that logical Heracliteanism, but more

---

[23]Note that according to the Aristotelian argument, the option of describing a language in which all contradictions are true but none of their components is true is open to Heracliteans, provided that in its metalanguage either not all contradictions are true or some of the components of a contradiction are true separately from it.

[24]That PNC might hold only for restricted versions has been pointed out several times by scholars. Substances and "essential predicates" are the usual candidates for which oPNC certainly holds (cf. [11, 17, 19]), although sometimes a slippery slope from the restricted oPNC about essential predicates to a general oPNC about any predicates is attempted; see for example [33]. Tahko [32] defends that oPNC holds certainly for concreta and its properties in "genuinely possible worlds with macroscopic objects", but not necessarily for every object of every possible world, not even for every object of every physically possible world. More recently, Coren [10] has recently provided an interpretation like ours, but attempting to show that Aristotle can succeed in defending a principle much stronger than MSPNC.

[25]See also [30, §19]; [31, Ch. VIII].

obstreperously since most of them, unlike Heracliteanism, are precisely not framed as foes of PNC. In order to demonstrate PNC and only PNC, these demonstrations would have to assume that only their premises —say, true first principles for Kant or laws of thought in Boole's case, let us denote them '$M_1, \ldots, M_n$'— but not also something excluded by them hold. If PNC, in the form of MSPNC, were not logically prior to any of those premises then it would also be possible to adopt Not-M_i for any of them, and then either proving Not-PNC, or both PNC and Not-PNC, but not only PNC, regardless the formulation each of them gives of PNC. Therefore, neither Kant's, nor Boole's nor Russell's would be demonstrations of PNC, at least not in the Aristotelian sense. The original sin of these demonstrations consists in most cases in mistaking self-evidence or obviousness by the Aristotelian technical *Certainty*, which are different as already mentioned in Sect. 4. Other principles might be more evident than PNC, but it does not mean that they are more certain and thus they cannot be used to demonstrate it.

As a step in his rebuttal of PNC, Priest says, more modestly than Kant, Boole or Russell, that "it is not clear that it is a first principle", that "it is not clear that the LNC cannot be demonstrated" and then he sketches a demonstration:

> Even granting this [Aristotle's] notion of demonstration, it is not clear that the LNC cannot be demonstrated, since it is not clear that it *is* a first principle. Consider, for example, the Law of Identity, $\Box(\alpha \to \alpha)$. Though nothing is completely uncontentious, there is hardly *any* disagreement about the correctness of this Law. (…) And given this, $\neg\Diamond(\alpha \wedge \neg\alpha)$ follows from one application of the rule of inference: $\Box(\alpha \to \beta) \vdash \neg\Diamond(\alpha \wedge \neg\beta)$. [New paragraph] One might object that the principle of inference here simply presupposes the LNC. This is moot: it is valid, for example, in the semantics of *In Contradiction*, chapter 6, according to which contradictions may be true. But whatever one says about this particular case, the point remains: it is not at all obvious that no proof of the LNC in Aristotle's sense is possible. ([26, 13, italics in the original])

Priest has attempted thus a demonstration of PNC using principles that seem more certain than PNC itself. One could demur at this point that Priest, like Lewis in his argument against the possibility of debate, uses a notion of certainty different from Aristotle's and that it could be that, on further, more detailed inspection, these principles show themselves as no more certain than PNC. But the standard Aristotelian move against both Detractors of PNC —like Heracliteans— and its Demonstrators —like Kant, Boole or Russell—, is that PNC has to hold for the premises of their arguments to only hold good and not also not to hold, so PNC is prior to the premises used whether to refute it or to demonstrate. However, Priest does not accept the priority of PNC in this sense and says that, for any of his premises or rules P_i, it is possible that P_i and Not-P_i. So even if PNC is not demonstrable under Not-(R1), Not-(R2), Not-(P1), Not-(P2) and Not-(P3), which he could endorse, it is under (R1), (R2), (P1), (P2) and (P3), which he also could endorse. That PNC could be both demonstrable and indemonstrable should worry only someone who already endorses PNC.[26] Thus, the Aristotelian move has seemingly left the friend of contradictions Priest unharmed.

[26]In [27, 241ff], Priest explicitly says that some arithmetical statements (in an inconsistent arithmetic) are both provable and not provable.

Priest could make the same move with Lewis' premises. But this easily generalizes. If PNC does not hold, for any proposition *A* it could be that *A* and not-*A*. There would be no purely logical way to say when a proposition does not hold and only does not hold, and then ruling it out (cf. [26, 107]). In particular, a demonstration (or the lack of one) of *A* would be at most a first step towards a verdict on its acceptability.

7 Methodologists and the Fruitfulness of the Debate

But that for Priest there is no logical way to rule things out does not mean for him that there is not a way at all:

> That a person may sometimes be able to accept a contradiction rationally, and that there is nothing in the domain of formal semantics ever to stop a person accepting a contradiction, I do not dispute. That a person can always accept a contradiction rationally is a blatant non sequitur, which I reject. It does not follow from the fact that some contradictions are rationally acceptable that all are, nor does it follow from the fact that there is nothing in formal semantics against it that it can be done rationally. ([27, 104])

For example, Priest thinks that *I am a frog* is a belief less rationally acceptable than *The Liar is both true and false* (cf. [26, 14]). According to him, the rationality of individual logical principles derives from the degree of rational preference or acceptance of the logical theory to which they belong, so these troubles with ruling out things can be solved by turning to the realm of pragmatics and the discussion of the rational acceptability of theories in general (cf. [26, 123]). Thus Priest counts finally not as a demonstrator but as a methodologist,[27] and given that he thinks that some theories in which PNC holds score worse in overall criteria for rational acceptability than those in which (allegedly) it does not hold, he is a methodologist detractor.

Let us grant that some dialetheists have managed to establish that *I'm a frog* has an extremely low degree of rationality and thus that it should be rejected. They cannot reject it by negating it, though, because according to them a proposition and its negation can both be true, and so acceptable. Nor they can reject it by saying that it is false, because for them a proposition can be false but also true, and so acceptable. But regardless how dialetheists can reject something, the important feature of acceptance and rejection is that they are incompatible for them, just like contradictories in many logics:

[27]Lewis mentions this:

> It is not an easy thing [to refute Priest's position, decisively and in accordance with customary rules of debate]. I myself think that it is an impossible thing: so much is called into question that debate will bog down into question-begging and deadlock. (On this point, Priest disagrees with me: he thinks that shared principles of methodology might provide enough common ground.) (Quoted in [27, xix].)

> Someone who rejects A cannot simultaneously accept it any more than a person can simultaneously catch a bus and miss it, or win a game of chess and lose it. If a person is asked whether or not A, he can of course say 'Yes and no'. However this does not show that he both accepts and rejects A. It means that he accepts both A and its negation. Moreover a person can alternate between accepting and rejecting a claim. He can also be undecided as to which to do. But do both he cannot. ([25, 618])[28]

But now Aristotelians may strike again in two respects, one regarding the appeal to epistemic virtues and another regarding the incompatibility between acceptance and rejection. First, the fierce supporter Lewis contented himself with agreeing to disagree with Priest, but Calm supporters, based on Aristotle's refutations, can argue that for every epistemic virtue E_i on the basis of which something is going to be excluded, assuming no version of PNC allows one to consider also *Not*-E_i, where it is something excluded by E_i. For example, if only *Simplicity* delivers simplicity (whatever that means) it is because it is not the case that both *Simplicity* and *Not-Simplicity* delivers simpler theories (whatever that could be) and so, theories more rationally acceptable. It is clear that Priest is in the sights of the Aristotelian now. A version of PNC has to hold at least at some level, contrary to the impression that for any proposition A, it could be that A and not-A, which led us to the realm of pragmatics: One wants to adopt an epistemic virtue E and not also *Not*-E, otherwise they would not play the role of ruling some propositions out as required by Priest.[29]

Second, and has already been pointed out by Berto ([4, 5]), for Priest there is a realm of notions (involving persons and buses, outcomes of games of chess, certain speech acts) that seem to fall under MSPNC, as is clear from what he says on acceptance and rejection. In Priest's dialetheism, rejection has to do the job of ruling some propositions, theories, claims, et cetera, out. But for that to work, he seemingly has to assume that the notions of acceptance and rejection stand in the relation sanctioned by MSPNC.[30]

What we have said about epistemic virtues would show that not even the well-intentioned proposal of other Methodologists like Bueno and Colyvan [8] to conduct debates about logical principles works so smoothly in their target case, PNC. They say:

> The crucial idea is that —similarly to what goes on in science— debates about logic typically involve a common core of assumptions that are shared by the various parties in the debate. This common core includes: (1) shared logical theories (that is, logical principles and rules), or (2) shared views about the aims of logic, or (3) shared methodological principles (broadly

[28] See also [26, 110].

[29] This reveals that Priest underestimated Aristotle's remark in *Posterior Analytics* mentioned in Sect. 4. Aristotle says that whenever one wants to demonstrate something but not also something excluded by it, one has to presuppose PNC. If it is presupposed in no demonstration, one always could demonstrate something but also something else excluded by what was supposedly demonstrated, hence all exclusions could be demonstrated. Given MSPNC, the fact that contradictory propositions can be accepted is irrelevant as to the *Indemonstrability* of (MS)PNC, although perhaps it counts against some form of *Universality* for especially stronger versions of it.

[30] To be fair, dialetheists and other detractors are right in questioning stronger forms of PNC, like "any sentence of the form $A \wedge \neg A$ is untrue in any valuation", and their assimilation to weaker forms like MSPNC.

> understood to include metalogical principles). Although usually items (1)–(3) are not all shared at the same time, *at least one of them* typically is. And as we will see, this provides enough common ground for debates about logic to be conducted and, in some cases, settled. ([8, 166])

Among their examples of "significant agreement" between parties there are certain affirmations, BC_m, which, they say, allow debate, but PNC would not be among them. For example, dialetheists and non-dialetheists agree in the following:
(BC1) The conjunction of true contradictions and classical logic implies trivialism.
(BC2) Trivialism is untenable.
Together with other implicit principles: What is untenable has to be rejected; if something is rejected then at least part of what implies it has to be rejected too, etc.[31] However, without some form of PNC, probably MSPNC, these basic affirmations shared by dialetheists and non-dialetheists do not rule out other affirmations that could entail a prima facie incompatibility with the alleged result of the agreement, and thus cannot do the job of providing either a justification or a rejection, and only a justification or a rejection, of PNC, and only of PNC.

Bueno and Colyvan [8, 170] suggest that the adoption of a principle not established through debate would undermine fallibilism, but fallibilism is surely an advisable stance. More generally, the backdrop of methodologism is, broadly speaking, the idea that logic is not fundamentally different from other scientific theories and that it is subject to the same desiderata and conditions as theories in natural science, as is clear from the first line of the above quotation and was also clear already in Priest's case. However, we do not think that the adoption of MSPNC suppose a great threat neither for fallibilism nor for the idea of logic as a theory as many others, though. Although MSPNC is flanked by a universal quantifier such that only trivialism escapes from it, there is a lot of room for fallibilism regarding what instances of the inner particular quantifiers (some properties…, some expressions…) should be accepted. One could claim that some properties and expressions fall under the scope of MSPNC and accept the claim using the usual methodology of natural sciences; also, and this was already pointed out by Berto [4, 185] in his own terminology, one can be wrong about such claims and rationally retract them again using the methodology of natural sciences. Then, neither fallibilism nor the idea of a methodology akin to that of the rest of science are threatened by the adoption of a principle like MSPNC.

The remarks above also help to explain the conflicting intuitions about the possibility of debate about PNC. MSPNC can be recognized as holding universally, as Defenders of PNC say; its outer universal quantifier can be taken as being neither demonstrable, for any demonstration of it would already require it, nor refutable, because it would be a true universal claim. But debate would be possible since MSPNC does not guarantee by itself its extension given the inner particular quantifiers. Moreover, MSPNC leaves room for some non-false or plainly true contradictions, as adversaries of stronger versions PNC have claimed. This debate would be fruitful because proposing and rejecting specific instances of MSPNC involve

[31] Cf. the reconstruction of the debates in [8, 170–173].

substantial claims. For example, even if dialetheists and classicists agree that acceptance and rejection fall under the scope of MSPNC, they still can fruitfully argue whether truth and falsity or specific contradictions also fall under it without calling into question MSPNC, that is, they can debate with enough common ground, as a fierce supporter like Lewis demanded.

8 Conclusions

In this paper we have investigated the possibility and fruitfulness of a debate on the principle of non-contradiction (PNC) by studying an Aristotelian approach to the current major stances on the issue. More in particular, we have argued that the Aristotelian refutative strategy, originally used in *Metaphysics* for dealing with detractors of PNC, can be used for a better understanding of the different standpoints involved in the contemporary debate on PNC.

After stating the version of PNC to be discussed, we have argued that one can find all the elements of "Calm supporters" of PNC already in Aristotle's works. A calm supporter holds that a debate on the PNC is both possible and fruitful, and that there is a version of PNC that even "Detractors" must endorse even if stronger forms of the principles might have counterexamples. Calm supporters, contrasting "Fierce supporters" like Lewis, think that a debate about PNC is both possible and fruitful because PNC and its properties need not be evident and can be honestly doubted or denied by detractors; and although PNC cannot be demonstrated and Detractors are doomed to fail in their quest for counterexamples, a debate with Detractors is possible because they can legitimately challenge the principle yet are subject to refutation. Refutation "silences" the denials by showing that they are ultimately committed to PNC.

We have also showed that this strategy of exhibiting an ultimate commitment to PNC serves for replying to "Demonstrators" and "Methodologists". Demonstrators deny the indemonstrability of PNC, but their alleged demonstrations are already committed to PNC in that their premises are intended to mean something and not also something incompatible with that, so they are not demonstrations. The mistake lies in most cases in confusing self-evidence with certainty. Other principles might be more evident than PNC, but it does not mean that they are more certain and thus they cannot be used to demonstrate it. A difficult case is Priest's demonstration of PNC, because at the same time he accepts the charge that it is committed to PNC, so it is both demonstrable and indemonstrable. He endorses then a methodologist approach according to which principles and logics incorporating them should be accepted or denied on the basis of the usual criteria for theory choice. But Methodologists fall short by the same reasons as detractors and demonstrators: They presuppose PNC since their methodological principles for theory choice are intended to mean something but not also something excluded by them; for example, if simplicity is an epistemic virtue, it is thought to exclude non-simple theories. However, the schematic character of PNC provides enough common ground for debate yet leaves room for disagreement and fallibilism.

Acknowledgements This work was supported by the PAPIIT project IA401117 “Philosophical Aspects of Contra-Classical Logics”. We would like to thank Daniel Cohnitz, Alex Davies, Víctor Cantero, Maite Ezcurdia, Eduardo García-Ramírez, Tuomas Tahko and Pedro Stepanenko, as well as the audience at the conference *Trends in Logic XVI*, for their comments on previous incarnations of this paper, some of them known as “Kantham Priest vs. Aristot Lewis”. The anonymous referees deserve special mention for their extremely useful reports.

References

1. Aristotle: Metaphysics. In *The Complete Works of Aristotle*, ed. J Barnes (vol. 2). Princeton N.J.: Princeton University Press, 1984.
2. Aristotle: Posterior Analytics, edited and translated with a commentary by Jonathan Barnes, Oxford: Clarendon Press, second edition, 1994.
3. Aristotle: Prior Analytics, Robert Smith (trans.), Indianapolis: Hackett, 1989.
4. Berto, Francesco. 2008. Adynaton and material exclusion. *Australasian Journal of Philosophy* 86 (2): 165–190.
5. Berto, Francesco. 2012. How to rule out things with words. In *New Waves in Philosophical Logic*, ed. Greg Restall, and Gillian Russell, 169–189. Great Britain: Palgrave Macmillan.
6. Boghossian, Paul. 2000. Knowledge of logic. In *New Essays on the A Priori*, ed. Paul Boghossian, and Christopher Peacocke, 229–254. New York: Oxford University Press.
7. Boole, George. 1854. *An Investigation of the Laws of Thought on which are founded the Mathematical Theories of Logic and Probabilities*. Cambridge: Cambridge University Press.
8. Bueno, Otávio, and Mark Colyvan. 2004. Logical non-apriorism and the "law" of non-contradiction. In *The Law of Non-Contradiction: New Philosophical Essays*, ed. Jc Beall, Graham Priest, and Bradley Armour-Garb, 156–175. New York: Oxford University Press.
9. Castagnoli, Luca. 2010. *Ancient Self-Refutation: The Logic and History of the Self-Refutation Argument from Democritus to Augustine*. Cambridge: Cambridge University Press.
10. Coren, Daniel. 2018. Why does Aristotle defend the principle of noncontradiction against its contrary? *Philosophical Forum* 49 (1): 39–59.
11. Cresswell, Maxwell. 2004. Non-contradiction and substantial predication. *Theoria* 70: 166–183.
12. Dancy, Russell M. 1975. *Sense and Contradiction: A Study in Aristotle*. Dordrecht: Reidel.
13. Dutilh Novaes, Catarina. 2012. *Formal Languages in Logic -A Philosophical and Cognitive Analysis*. Cambridge: Cambridge University Press.
14. Field, Hartry. 2000. Apriority as an evaluative notion. In *New Essays on the A Priori*, ed. Paul Boghossian, and Christopher Peacocke, 117–149. New York: Oxford University Press.
15. Graham, Daniel. 2006. *Explaining the Cosmos: The Ionian Tradition of Scientific Philosophy*. Cambridge: Cambridge University Press.
16. Grim, Patrick. 2004. What is a contradiction? In *The Law of Non-Contradiction: New Philosophical Essays*, ed. Jc Beall Graham Priest, and Bradley Armour-Garb, 49–72. New York: Oxford University Press.
17. Irwin, Terence. 2006. *Aristotle’s First Principles*. Oxford: Oxford University Press.
18. Kant, Immanuel. 1755. Concerning the principle of contradiction. In *Kant. Theoretical Philosophy 1755-1770*, ed. David Walford, and Ralf Meerbote, 6–10. Cambridge: Cambridge University Press.
19. Kirwan, Christopher. 1993. *Aristotle, Metaphysics (Books Γ, Δ, E)*, 2nd ed. Oxford: Oxford University Press.

20. Lear, Jonathan. 1988. *Aristotle: The Desire to Understand*. Cambridge: Cambridge University Press.
21. Lewis, David. 1982. Logic for equivocators. *Noûs* 16 (3): 431–441; Reprinted as Chapter 7 of Lewis. 1998. *Papers in Philosophical Logic*, 97–110, Cambridge: Cambridge University Press.
22. Lewis, David. 2004. Letters to Beall and Priest. In *The Law of Non-Contradiction: New Philosophical Essays*, ed. Jc Beall Graham Priest, and Bradley Armour-Garb, 176–177. New York: Oxford University Press.
23. Ole Thomassen Hjortland. 2017. Anti-exceptionalism about logic. *Philosophical Studies* 174 (3): 631–658.
24. Plato: The Republic, Desmond Lee (trans.), Penguin Classics, 1955.
25. Priest, Graham. 1989. Reductio ad absurdum et modus tollendo ponens. In *Paraconsistent Logic. Essays on the Inconsistent*, ed. Richard Routley, Graham Priest, and Jean Norman, 613–626. München: Philosophia Verlag.
26. Priest, Graham. 2006. *Doubt Truth to be a Liar*. Oxford: Oxford University Press.
27. Priest, Graham. 2006. *In Contradiction*, 2nd ed. Oxford: Oxford Clarendon Press.
28. Priest, Graham. 2007. Paraconsistency and dialetheism. In *The Many Valued and Nonmonotonic Turn in Logic*, vol. 8, ed. Dov Gabbay, and John Woods, 129–204., Handbook of the History of Logic The Netherlands: Elsevier.
29. Rescher, Nicholas, and Robert Brandom. 1980. *The Logic of Inconsistency: A Study in Non-Standard Possible Worlds Semantics and Ontology*. Oxford: Basil Blackwell.
30. Russell, Bertrand. 1903. *The Principles of Mathematics*. London and New York: Routledge.
31. Russell, Bertrand. 1912. *The Problems of Philosophy*. New York and Oxford: Oxford University Press.
32. Tahko, Tuomas E. 2014. The metaphysical interpretation of logical truth. In *The Metaphysics of Logic: Logical Realism, Logical Anti-Realism and All Things In Between*, ed. Penelope Rush, 233–248. Cambridge: Cambridge University Press.
33. Wedin, Michael V. 2000. Some logical problems in metaphysics gamma. *Oxford Studies in Ancient Philosophy* 19: 113–162.
34. Wedin, Michael V. 2014. The science and axioms of being. In *A Companion to Aristotle*, ed. Georgios Anagnostopoulos, 123–143. Massachusetts: Wiley-Blackwell.
35. Whitaker, Charles W. A. 1996. *Aristotle's De Interpretatione: Contradiction and Dialectic*. Oxford: Oxford Clarendon Press.
36. Whitehead, Alfred North, and Bertrand Russell. 1910. *Principia Mathematica*, vol. I, 2nd ed. Cambridge: Cambridge University Press.
37. Williamson, Timothy. 2013. Logic, metalogic and neutrality. *Erkenntnis* 79 (2): 211–231.
38. Williamson, Timothy. 2017. Semantic paradoxes and abductive methodology. In *Reflections on the Liar*, ed. Bradley Armour-Garb, 325–346. Oxford: Oxford University Press.

Keeping Globally Inconsistent Scientific Theories Locally Consistent

Michèle Friend and Maria del Rosario Martínez-Ordaz

Abstract Most scientific theories are globally inconsistent. *Chunk and Permeate* is a method of rational reconstruction that can be used to separate, and identify, locally consistent chunks of reasoning or explanation. This then allows us to justify reasoning in a globally inconsistent theory. We extend chunk and permeate by adding a visually transparent way of guiding the individuation of chunks and deciding on what information permeates from one chunk to the next. The visual representation is in the form of bundle diagrams. We then extend the bundle diagrams to include not only reasoning in the presence of inconsistent information or reasoning in the logical sense of deriving a conclusion from premises, but more generally reasoning in the sense of trying to understand a phenomenon in science. This extends the use of the bundle diagrams in terms of the base space and the fibres. We then apply this to a case in physics, that of understanding binding energies in the nucleus of an atom using together inconsistent models: the liquid drop model and the shell model. We draw some philosophical conclusions concerning scientific reasoning, paraconsistent reasoning, the role of logic in science and the unity of science.

The original version of the chapter was revised: The name of the author has beeen corrected from "First Name: María and Family Name: del Rosario Martínez-Ordaz" to "First Name: Maria del Rosario and Family Name: Martínez-Ordaz". The correction to the chapter is available at https://doi.org/10.1007/978-3-319-98797-2_15

M. Friend (✉)
Department of Philosophy, The George Washington University,
801 22nd. St. N.W., Washington, DC 20052, USA
e-mail: michele@gwu.edu

M. R. Martínez-Ordaz
Instituto de Investigaciones Filosóficas, Universidad Nacional Autónoma de México,
Circuito Maestro Mario de la Cueva s/n, Ciudad Universitaria, 04510 Coyoacán,
Ciudad de Mexico, Mexico
e-mail: martinezordazm@gmail.com

W. Carnielli and J. Malinowski (eds.), *Contradictions, from Consistency to Inconsistency*, Trends in Logic 47, https://doi.org/10.1007/978-3-319-98797-2_4

1 Introduction

Inconsistency toleration is practiced when we recognise sensible reasoning from inconsistent information. Moreover, under inconsistency toleration, such reasoning is not considered to be a threat to rationality. It is claimed that inconsistency toleration is quite common in science. Moreover, from studying the history of science we learn that, at some time in their development, most scientific theories have been inconsistent ([25, 27, 34, 44]). Some of the most famous examples of this are: Aristotle's theory of motion ([46]), the early calculus ([5, 25, 46]), Bohr's theory of the atom ([11, 17, 25]), and Classical Electrodynamics ([19, 20]), among others.

Despite the historical examples, accepting the inconsistent character of science meets with resistance. Whether the resistant are explicitly aware of it or not, the reason for their resistance rests on the deeply entrenched view that the world is a unified and consistent whole. It follows that our true scientific theories must also be consistent in order to reflect this view of the world. We can trace the origins, at least in the West,[1] to Aristotle's denial of contradictions in the world and of his question begging ([44], chapter 1) arguments that *ex contradictione quodlibet* (from a contradiction any formula, proposition or sentence we can write in the language follows) is a valid argument form. Thus, from Aristotle we learn that reasoning from a contradiction is somehow irrational, false or wrong.

Today, the view is reflected in the underlying logic used in making arguments in science. The more widely accepted formal representations of such reasoning are found in classical or constructive logic (where *ex contradictione quodlibet* arguments are valid).[2] If the basic principles of classical or constructive logic[3] are correct then, "an inconsistent theory implies any conceivable observational prediction as well as its negation and thus tells us nothing about the world" ([24]: 79). Reasoning under inconsistency is called explosive reasoning since everything we can express in the language is true.

In order to reconcile the present and historical cases of alleged inconsistent science with the repugnance of explosive reasoning, logicians have developed reasoning strategies that model sensible reasoning from inconsistencies without arriving at arbitrary conclusions [4, 7, 10, 11, 47]. Formal representations of reasoning where *ex contradictione quodlibet* is invalid include the relevant and paraconsistent theories. One of the common strategies of the formal theories is to separate the original inconsistent set of formulas or sentences into consistent subsets. Arguably, this seems

[1]Of course, we do not mean the 'West' in the geographical sense. We mean it in the sense of a cultural tradition in science.

[2]Not all scientists have a thorough training in formal logic, so they might not be acquainted with the terms 'classical logic' or 'constructive logic', nevertheless, these are accepted as the cannons of reasoning and rationality by those who do have some training in logic, and their authority is accepted, for reasons of division of labour. It is less fashionable now for scientists to study formal logic. It was more common in the past.

[3]Classical and constructive approaches or reasoning styles are the most common in mathematical and scientific practice. We shall discuss an important limitation of the formal representation of such reasoning in the next section.

to be a fruitful, intuitive and common strategy when reasoning with inconsistent information [4].

Such separation exercises allow us to explain how we reason in the majority of the historical cases of inconsistent science. We say that from the *global* perspective, the scientific theories are inconsistent but that they have several important *local* consistent sub-parts. This also explains how it is possible to apply inconsistent science to technology, to wit:

> A nice cutting edge high-tech example of such integration is the global positioning system (GPS): by means of satellites kept in place by Newtonian physics, and atomic clocks ruled by quantum mechanics and corrected by special and general relativity, this system maps the spherical surface of the round earth on a geocentric grid (or rather, a geostatic grid), and gives advice to people on the ground from a flat-earth point of view. ([12], 266).

The theories, say of the GPS system, have consistent sub-theories: the application of Newtonian mechanics to solve one problem, the application of quantum mechanics to run the clocks, the application of the relativity theories to correct the clocks for their interaction with clocks on earth and so on. These theories of application are more specialised theories, and these sub-theories are pairwise inconsistent with each other. Taking this into account, the phenomena of inconsistency toleration has been studied by logicians by distinguishing *global* analyses and *local* analyses of the alleged inconsistent theories.[4]

One of the strategies for modelling non-explosive inconsistent reasoning that makes use of separation into local consistent sub-theories is *Chunk and Permeate* ([4, 10, 11]; Priest 2015; [7]). The underlying strategy is to separate a given inconsistent set of sentences into consistent subsets and cordon them off by calling the subset a 'chunk'. We then let only *some* information permeate from one chunk to the next. This strategy has been used to model some of the most popular examples of inconsistent science such as the early calculus (cf. [10]; Sweeney 2014), Bohr's Hydrogen Atom (see [11]), the Dirac Delta function (cf. [4]) and Lobachevsky's model of hyperbolic geometry for indefinite integrals ([18], 162–172).

While the chunk and permeate strategy has been successfully used to model particular examples of inconsistent scientific reasoning, it still lacks a *systematic* method for separating chunks and for evaluating such separations for consistency. Here we provide an almost effective method[5] for separating inconsistent sets of formulas and sentences. Because the method is not quite effective, we can only claim to make the modelling *more* systematic, not systematic in the sense of giving an effective procedure, although in some restricted cases we conjecture that it is possible to give an effective procedure. This is reserved for future work.

---

[4]Here we shall focus only on global and local analyses of (in)consistency regarding specific bodies of knowledge. In Sect. 3 we shall introduce the notions of global and local consistency and their representation using bundle diagrams. In Sect. 6.2 we shall briefly introduce the corresponding characterisations of *global (in)consistency* and *local consistency* when dealing with inconsistencies in empirical sciences.

[5]It is not completely effective, and can only be made so under quite rigid and formal circumstances. See the conclusion for more details about these limitations. Nevertheless, even as an almost effective method, or even as a heuristic in informal reasoning, it will be quite useful.

The significance of our work is that we can more clearly discuss some philosophical considerations on the subject of inconsistency toleration in formal and empirical sciences. More specifically, in this paper we focus on three important issues: the role of liar cycles and inconsistent reasoning in the formal sciences, the use of Chunk and Permeate and Bundled Chunk and Permeate for modelling, reconstructing, explaining and providing understanding of inconsistent scientific reasoning, and we discuss the implications of inconsistency toleration in science especially in the light of the research programmes that aim at the unification of science.

In order to address these issues, the paper is divided in two main parts: the first one is devoted to logic and mathematics. We elaborate on the benefits of using Chunk and Permeate when modelling inconsistent scientific reasoning, and present our method. This part includes Sects. 2–5. In particular, in Sect. 2 we explain why it is that reasoning with inconsistencies is assumed to be problematic from the point of view of the philosophy of logic. In Sect. 3 we introduce the Chunk and Permeate strategy and give reasons for looking at Bundled Chunk and Permeate. In Sect. 4 we introduce bundle diagrams. In Sect. 5 we use the bundle diagrams to individuate chunks and determine permeation for liar cycles.

In the second part of the paper, we discuss science. This part includes Sects. 6–10. In Sect. 6, we make some general remarks about the problem of inconsistency in science. In Sect. 7, we introduce some considerations from the philosophy of science about how to individuate and combine mutually inconsistent theories or models. In Sect. 8 we provide an example of a globally inconsistent union of models and we apply Bundled Chunk and Permeate, to this example. In Sect. 9 we present further insights concerning why inconsistent groups of theories are thought to be a problem for science. Finally, in Sect. 10 we draw some philosophical conclusions concerning the unification of science, the nature of consistency and reasoning paraconsistently, and what this means, while using only consistent formal representations of logical reasoning locally.

2 Trivialism and Modern Mathematics

We shall be looking at the problems with trivialism and modern mathematics through two lenses, the classical lens and the constructivist lens. We begin with trivialism and why it is a problem.

Under classical, model-theoretic conceptions of semantics, a trivial theory is one where every formula in the language is true; under constructivist or proof-theoretic conceptions a trivial theory is one where any well-formed formula of the language can be derived. A trivial arithmetical theory would have it that $2 + 2 = 4$, but also $2 + 2 = 19$, $2 \times 93 = 6$ and so on. This is not a useful theory of arithmetic for science. In fact, it is a disastrous theory of arithmetic, since it is completely undiscerning between the true theorems or equations and the false ones (as seen from a more traditional consistent and classical conception of arithmetic). There is no false statement, only ungrammatical ones, and ungrammatical statements are, arguably, not counted as statements. Grammatical statements of the theory are all true, all derivable and their negations are all true and derivable. The conceptions of arithmetic error and correction are lost, and arguably ([44] Ch. 3), meaning is also lost. Trivial theories are to be avoided according to the more common present practice of mathematics.[6]

Most mathematicians claim[7] that they are classical or constructivist reasoners, ([23], 64–70) so they think that if there is a contradiction in their theory, then their theory becomes trivial.[8] In other words, one route to trivialism starts from classical or constructivist reasoning, you then meet a convincing contradiction that you do not think can be explained away, reason as you would through the *ex contradictione quodlibet* argument and you find yourself in a recognisably trivial theory. By *modus tollens*, if we think we are not in a trivial setting, since this might be thought to be *a priori* impossible, or we think that our theory is not trivial and we are unwilling to give up our classical or constructivist reasoning, then our theory had better not have any contradictions. So, we can avoid trivialism by remaining convinced that whatever looks like a contradiction in our theory must *a priori* not be one. This leap back from the brink of trivialism is quite common in science, and it explains, or excuses, the separated reasoning where the global theory is inconsistent but the local

[6]We thank an anonymous reviewer for reminding us that there are supporters of trivialism, and those who think that we can reason sensibly even in a trivial setting. In fact, this is almost what we show using Chunk and Permeate. 'Almost' means that there is some ambiguity as to what this means. See some remarks in the conclusion for elaboration. Since our concern is with present-day practice and reasoning in science, we maintain that at present there are no trivialists in science.

[7]Some relevant and paraconsistent logicians claim that such mathematicians are actually, as manifested in their reasoning behaviour, relevant or paraconsistent reasoners. This nuance will be addressed in the conclusion.

[8]The logicians and mathematicians who disagree with this, who think that *ex contradictione quodlibet* proofs are *invalid*, are relevant logicians or paraconsistent logicians. The *philosophical* difference is that relevant logicians insist on there being a relevant connection between premises and conclusion, paraconsistent logicians think that we can reason coherently with contradictions, or through a contradiction, and they model such reasoning. Briefly, in a paraconsistent logic, while you can derive an infinite number of formulas, as you can from any formula in any logic with a minimum set of inference rules, you cannot derive very much of interest from a contradiction. It is treated as a logical singularity. From $p \wedge \neg p$, you can derive p, $\neg p$, by $\wedge$-elimination, then by $\wedge$-introduction, you can derive $(p \wedge \neg p) \wedge p$ and so on, with double negation introduction you could derive $\neg\neg p$... The point is that you cannot get to an arbitrary q.

Logically, what distinguishes relevant from other paraconsistent logics is that relevant logicians, as part of the bigger substructural tradition, restrict some structural rules rather than operational ones. Non-relevant paraconsistent logicians change the behavior of the connectives (especially negation) while preserving the full set of structural rules of the language. This guarantees that they stay as close to classical logic as possible (i.e. Priest's LP).

This second way of putting the distinction reveals an important bias in this paper and for Chunk and Permeate in general: it appeals to a specific kind of logician/mathematician/scientist. Martínez-Ordaz would say that this particular kind of reasoner is one who admits that classical logic is along the right lines and is a good starting point and possibly thinks that formal representations of relevant reasoning sacrifice too much or change the reasoning too much. Chunk and Permeate then appeals to: classical, constructive and some (non-relevant) paraconsistent reasoners (those who think that inconsistency toleration is alright but we should nevertheless reason as consistently as possible).

pieces are consistent. When we meet an apparent contradiction in our classical or constructivist setting, we stay short of going through the *ex contradictione quodlibet* reasoning.[9]

Let us turn to modern mathematics. Modern Western mathematicians make proofs.[10] Often, these are only partly formal, so we might not notice a contradiction. Some proofs include sets of premises, lemmas or theorems that *belong to* theories that are inconsistent with each other, in the sense that they use information from different theories, and the theories themselves contradict each other. Some proofs include sets of premises, lemmas or theorems that are inconsistent with each other in the stronger sense that it is possible to derive a contradiction from them. Even worse, few mathematicians seem perturbed by this despite the threat of trivialism. How do we explain the lack of concern?

First note that none[11] of these proofs use an *ex contradictione quodlibet* proof or sub-proof, since this would bring disaster. In order to explain this, we might speculate that they are using a paraconsistent logic, or are reasoning paraconsistently, unbeknownst to them.[12] This is not an idle thought, since some paraconsistent logicians make this claim. If we agree with it, then it makes sense to use a paraconsistent logic to reconstruct the reasoning. But this would be disingenuous towards the claims, beliefs and practices of present day working mathematicians, since few of them claim to be, or believe that they are, reasoning paraconsistently, and they are qualified to make that judgement, at least *prima facie*.

We introduce Chunk and Permeate as a reconstruction of reasoning in the presence of contradiction that respects the claims, beliefs and practices of present day working mathematicians.

---

[9] Later in the paper, we shall see that this is exactly what Abramsky recommends.

[10] Not all mathematicians at all times finished their work with proofs. In the past, before the twentieth century in Europe, and in the colonies of the European countries, it became wide-spread in the institution of mathematics that results and ideas had to be proved. This is not the case in every mathematical culture, and it has not always been the case in European-based cultures. This is despite the fact that when detailed proofs were given, the proofs in Euclid set the standard for rigour of proof.

[11] The 'none' is meant as a challenge. The authors know of none that has been published, but of course some might have slipped into the published cannon.

[12] The difference is this: if they are using a paraconsistent logic, then they have recourse to a formal representation of the reasoning in the proof. If they are 'reasoning paraconsistently' then this is a looser notion. They are reasoning in such a way as to entertain and recognise contradictions but avoid trivialism. Here is the rub: which formal theory best represents their reasoning is usually ambiguous. Their reasoning is represented by a class of formal theories. They are reasoning in the spirit of paraconsistent reasoning in the sense of exercising damage control on the inconsistency.

3 Chunk and Permeate and General Remarks on Extending it

Let us highlight the original aim of [10, 11] in developing Chunk and Permeate (henceforth *C&P*). It was to reconcile the fact that sometimes mathematicians and scientists reason with inconsistent premises with the fact that they deny that this is possible or makes any sense. While it may seem sensible to those used to paraconsistent reasoning to argue that the inference procedures of such mathematicians should be represented by a paraconsistent formal logic, it is not always clear that the underlying logic is represented by any of the standard formal representations of paraconsistent reasoning—or that it can be formally represented at all ([10]: 379).[13]

The *C&P* strategy consists in dividing a given a proof with inconsistent premises into consistent subsets, called 'chunks', and to only allow some information to permeate from one chunk to the next.[14] It is assumed that within each chunk we have perfectly 'acceptable' (i.e. consistent) reasoning that can be represented using a classical or constructive formal logic.

Sharing Brown and Priest's original intention, our purpose is to formally depict only classical or constructive reasoning (within chunks) in cases where the premises are inconsistent with each other. This restricts the more general method of *C&P* because there is nothing forbidding us *a priori* from using a paraconsistent logic within a chunk or letting formal representations of paraconsistent conceptions to permeate from one chunk to another. We set aside such possibilities here because we are holding ourselves to the more *wide-spread* current standards in mathematical proofs.

Following the *C&P* strategy, we distinguish between two different types of chunk: source chunks and target chunks. The former are the input chunks, the ones that contain the original information that is often mixed in mathematical reasoning, while the latter are the output chunks, the ones that contain the desired results of the proofs [10]. Between chunks, we only allow to permeate the information we need to reach the conclusion of the chunk. We begin with the source chunks and end with the target chunk.

We should mention that a proof reconstructed with *C&P* loses cut-elimination, in the sense that premises are not always available in any chunk in the re-construction,

---

[13] Even though, in principle, one could also have relevant or paraconsistent reasoning within a chunk, we ignore this possibility here out of the respect for the prevailing claims beliefs and practices of working mathematicians. See Priest (2015) for an example of paraconsistent logics within chunks.

[14] An interesting question is whether we can use the chunk and permeate strategy on an *ex contradictione quodlibet* proof. Of course we can, in two different ways: one is to preserve classical validity, so the proof just is a demonstration that anything (written correctly in the formal language) can be derived from inconsistent premises. So the whole proof is one chunk. The second way is to separate the negated *reductio* inference from the double negation elimination, thus 'preserving' consistency within each chunk. A negated *reductio* inference is one where we conclude the negation of the hypothesis as opposed to the opposite of the hypothesis. If we hypothesise 'q', and this leads to a contradiction, then we conclude the negation (and opposite) '$\neg q$'. If we hypothesise '$\neg q$', we would conclude the negation, (and not the opposite) '$\neg\neg q$'.

[7]b. Premises have to be present in a chunk to be consulted in a chunk-sub-proof. Classical and constructive reasoning places no such restriction on the use of premises. For this reason, *C&P* proofs are non-classical and non-constructive. However, we need not be alarmed. There are many formal systems of proof where cut-elimination is absent; but more important, the loss of cut-elimination *almost passes unnoticed* in each *particular C&P* proof. For, we might prove cut-elimination in a chunk, or use cut-elimination within a chunk. It is only in the overall strategy of the proof that we lose cut elimination. Put another way, under the *C&P* strategy, given some premises, especially inconsistent ones, we do not countenance the closure of all inferences from the premises since this would be the trivial theory in that language. Since the mathematicians themselves do not consciously avail themselves of the trivial theory, we think it is legitimate to model their practice using *C&P*. Making note of this just makes explicit some of the philosophical subtleties involved in reasoning in ways that are closer to reasoning 'paraconsistenly' while not having a particular formal representation of paraconsistent reasoning in mind.

We extend *C&P* to model scientific *understanding* and problem solving, not just reasoning and making arguments, and we give a more rigorous *characterization* of chunks.[15] In past reconstructions, choosing the chunks was largely a matter of feel, with hints taken from the original proof. The more rigorous characterization we propose here is meted out in terms of bundle diagrams, but it could also be done more rigorously in terms of cohomology theory and sheaf theory [1] or in terms of a *pivotal consequence relation*.[16]

The notion of a pivotal consequence relation is used to maximize sets of assumptions or axioms or rules of inference, up to cut elimination. This would be a way of distinguishing chunks from each other. These extensions have not yet been worked out for *C&P* explicitly.[17] However, all of these more precise, rigorous, systematic and formal approaches to defining chunks and the permeability relation might *suffer* from being too precise because they would also have to be adapted to general understanding as opposed to reasoning or deducing, and worse, they might be applicable only in certain sorts of proof – those that can be expressed in the respective formal languages. The pivotal consequence relation concept coupled with maximal sets of assumptions up to cut-elimination is limited to cases that we can express in propositions and in terms of clear and explicit rules of inference. Extending the *C&P* strategy using cohomology theory or sheaf theory might also be too precise for the purposes of reconstructing some of the reasoning in *science*, although Abramsky

[15]It would be nice to make these *maximal*, but to prove that they are might not be possible. Similarly, to give a method for checking for maximal chunks might not be possible. There might be two *C&P* reconstructions that have the same number, or size, of chunks.

[16]See the work of ([30, 31]) for the introduction of this concept and Piazza and Pulcini [41] for the notion of finding the maximal set of assumptions that could then be used, again, to extend the *C&P* strategy by using the maximal set of assumptions to define a chunk.

[17]It would make a nice future project to look into the possibility of more rigorously defining the chunks in this way. Moreover, there promises to be some clean ways of working out what information permeates using the definition of complementary sequent and complementary system. See Piazza and Pulcini (2016) for details. We thank Pulcini for the suggestion in private correspondence.

et. al. do this for quantum mechanics, but without *C&P*. While this might work for highly mathematical areas of physics, it would be too precise in cases where we find it difficult to fit the concepts of science to the concepts and language of cohomology theory or sheaf theory. The scientific concepts might not be ready (yet) to be represented in this way. On the other hand, if the scientific theory is amenable to such representation, then it might be quite revealing to work through the *C&P* exercise. Generally, the more logical, formal or mathematical a science is, the more amenable it is to a more rigorous extension of *C&P*.

The bundles that we introduce in the next section are quite flexible and can be thought of in several very different ways. They are suited to representing scientists' more general understanding and reasoning as well as representing proofs. They are more flexible, but when combined with *C&P* give fairly rigorous guidance for individuating chunks. Thus, Bundled *C&P* takes us a step beyond the existing guidelines which are to individuate chunks by 'trying to follow the original intentions of the author of the proof'. They take us a step towards greater rigour. In particular, the further step would be to use cohmology theory, sheaf theory, and pivotal consequence relation approaches to individuating chunks. We believe that the approach that we introduce here, will sharpen our understanding of *C&P* as well as our understanding of the scientific practice when dealing with inconsistency by separating information according to context or background theory.

As we can already see, there are both practical and conceptual limitations to our extension of the method. We shall discuss some of them further in the conclusion.

4 Bundles: Local and Global Consistency

We are interested in inconsistencies. In particular, in inconsistencies in information being used to reason or understand phenomena in mathematics or science. We are interested in cases that are a little sophisticated: where we do not simply have a formula or sentence as one piece of information and the negation or denial of the (otherwise) same formula or sentence.

In order to introduce the bundles, we follow Abramsky et. al. and focus on the re-enforced liar paradox, also called 'liar cycles', where one person says of a second that everything he says is true, while the second says of the first that everything he says is false. This is a liar cycle of two. There can be liar cycles of three, four and so on, and because they might be extensive, we might not be certain whether we are in a liar cycle or not because the cycle is too large, or of indeterminate size. This makes the inconsistency more sophisticated.

There is a similar situation with some proofs in mathematics and computer science. In the langage of classical logic: creating a model or making a derivation can influence what other models are then possible. We add more information—a new result from another theory—and the models that satisfy this new information might preclude the first models—this is a cycle of two. The model cycles might be larger, up to indefinitely large. We might not know that satisfying some premises with a class of models precludes our satisfying other premises with the same models.

In the language of proof-theory,[18] or of constructive logic, deriving a certain theorem might set parameters on what can be derived next, and further down the line. In an informal proof, we might find that by ignoring some of the work we did earlier, we derive something that contradicts what we derived earlier. This is only possible if we are reasoning under suppositions or hypotheses, and the suppositions or hypotheses are important just for a sub-proof. We might not execute the derivation needed to see the contradiction, and so not be aware of the contradictory *milieu* we are in.

The other places where we see such reasoning is in quantum mechanics, reasoning from inconsistent data sets and so on ([1], 1). Or, there are situations where a mathematician borrows theorems or results from various theories to prove her conclusion, suspects that she might be flirting with inconsistency, but is, nevertheless, confident (or the mathematical community is confident) that her result stands. For example, there might not be a tight and loyal translation between the theories, and usually not even an equi-consistency proof between the theories. The mathematician then borrows information from other theories that is locally consistent. But if we were to mix all the theories together, we might well be able to derive a contradiction. Moreover, she thinks she is reasoning classically or constructively. Such reasoning is sensitive to context: that it should be local. For this reason, Abramsky et. al. call this 'contextual' reasoning. ([1], 1) When we develop Bundled *C&P*, contextual reasoning will be treated as a chunk.

We can represent liar cycles, and similar sorts of reasoning using bundle diagrams from topology. Precisely:

> The key idea is to understand contextuality as arising where we have a family of data which is *locally consistent but, globally inconsistent*. This can be understood and very effectively visualised... in topological terms: we have a base space of *contexts* (typically sets of variables that can be measured or observed), a space of data or observations fibered over this space, and a family of local sections (typically valuations of the variables in the context) in these fibres. This data is consistent locally but not globally: there is no *global* section defined on all the variables that reconciles [makes together consistent] all the local data. In topological language we say that the space is "twisted" and hence provides an *obstruction* to forming a global section. (Our emphasis, [1], 1).

In Sect. 5, we shall extend the bundle diagrams to accommodate other cases, by considering other sorts of variables (base spaces) and other sorts of valuations on those variables. We shall then see how they fit with *C&P* to vindicate mathematician's practice of reasoning with inconsistent premises.

Bundles are a type of diagrammatical representation. We shall first construct a simple diagram, showing a consistent set of formulas (Figs. 2, 3, 4), then we shall show liar cycles of three (Figs. 5, 6). We shall then widen the cycle to five (Fig. 7). Next, we change some of the parameters on the bundle diagram to accommodate different sorts of proof; and finally, we transpose this idea to the notion of *C&P*

[18] If we are doing formal proof theory, then there is no danger of inconsistency. However, here we are thinking in terms of informal proofs or proofs using suppositions. We move from the model theory story to the proof theory story to respect classical reasoning and constructive reasoning, respectively.

as a methodologically tight rational reconstruction of reasoning with inconsistent premises, solve problems or trying to understand scientific phenomena from the point of view of scientific theories that contradict each other.

We introduce the bundle diagrams. An easy bundle diagram for a consistent set of formulas consists in the following. Assuming that everything Aristotle says, Plato says and Socrates says is internally consistent, we make up our *base space* of: (A) everything Aristotle says, (B) everything Plato says and (C) everything Socrates says. Represent this with three points on a horizontal surface. Rising vertically upwards from the points (A), (B) and (C), we draw fibres. See Fig. 1.

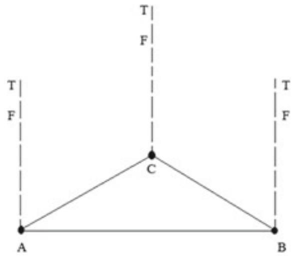


Fig. 1 Base space

Along each fibre are two stops: the possible valuations (true or false) of the set of sentences uttered by Aristotle, Plato and Socrates respectively. Let the lower stops represent false and the upper stops represent true. That is, it is possible that everything Aristotle says is true, and it is possible that everything he says is false, similarly for Plato and Socrates.

We add further information. Each of Aristotle, Plato and Socrates utter a special sentence. Aristotle says: everything Plato says is true. Plato says: everything Socrates says is true, and Socrates says: everything Aristotle says is true. Assuming that what all three say is true, this connection is represented by drawing an edge from the T stop, up the fibre from Aristotle, to the T stop, up the fibre from Plato and drawing an edge from the T stop, up the fibre from Plato, to the T stop up the fibre from Socrates. Finally draw an edge from the T stop, up the fibre from Socrates, to the T stop up the fibre from Aristotle. See Fig. 2. Drawing these edges makes a 'section'.

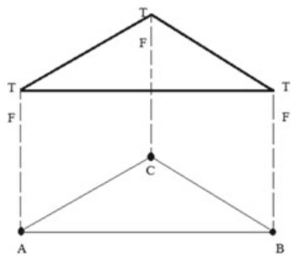


Fig. 2 Bundle diagram global consistency (all true)

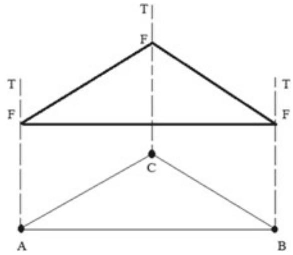


Fig. 3 Bundle diagram global consistency (all false)

This set of three edges represents the idea that Aristotle, Plato and Socrates say only truths, and that they attribute truth to each other (we ignore the direction of the attribution and the historical veracity of the claims in the base space). More technically, the three edges constitute a *closed path* that *traverses* each of the *fibres* only once ([1], 7). This represents *global consistency* in what Aristotle, Plato and Socrates say. They could also have all said only falsehoods rather than truths, and attributed falsehood to everything each other says. This could still be a consistent set of sentences. In this case we would have a path connecting each of the F stops up the fibres. See Fig. 3. Global consistency can also occur with a mixture of truths and falsehoods. For example, see Fig. 4.

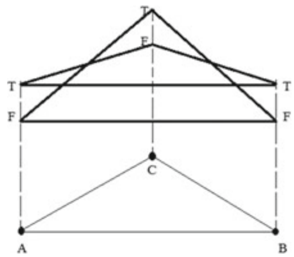


Fig. 4 Bundle diagram: global assignments

Aristotle might say that everything Plato says is true, but Plato says that everything Socrates says is false, and Socrates says that everything Aristotle says is false, but what he says is false, so the edge goes from the F stop up the fibre from Socrates to the T stop up the fibre from Aristotle. This is globally consistent, so we have a closed path that traverses all the fibres only once. But say Aristotle is uttering a falsehood when he says that everything Plato says is true. Then we have another closed path. See Fig. 4. There are all together eight possible paths traversing each of the fibres only once when we have a base space of three and two values up the fibres. Such a “closed path” is also called a “global assignment” ([1], p. 7). To introduce more vocabulary: any such closed path (traversing each fibre only once) is also called *univocal* since it assigns one value to each variable.

We do not always have global consistency, although we might have local consistency. Let us now consider a liar cycle of three. We have the same base space. The special sentences are the same, with one exception. This time, Aristotle says that everything Plato says is false. The bundle diagram now is given in Fig. 5. If we follow the path made by the edges, starting with assuming that what Aristotle says is true, it will cross the fibres twice.

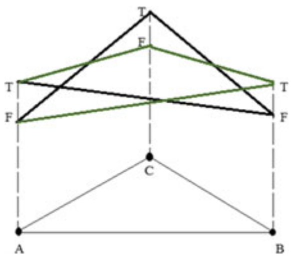


Fig. 5 Bundle diagram: liar cycle of three

If we start with the assumption that everything Aristotle says is false, then we have another path, that also crosses all of the fibres *twice*. See Fig. 6.

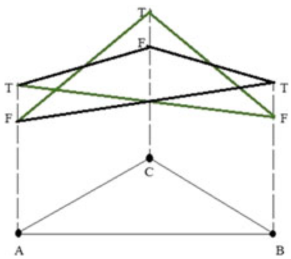


Fig. 6 Bundle diagram: liar cycle of three

This is because the variables: (A) what Aristotle says, (B) what Plato says, and (C) what Socrates says cannot be globally true, although any pair is locally, or pairwise, consistent. This is a liar cycle of three. We can expand it to four, five or more. We add Parmenides (D) and Heraclitus (E) between Plato and Aristotle. Plato's special sentence is now "Everything Parmenides says is true", Parmenides' special sentence is "Everything Heraclitus says is true", and Heraclitus' special sentence is "Everything Plato says is false". So this increases the liar cycle. See Fig. 7.

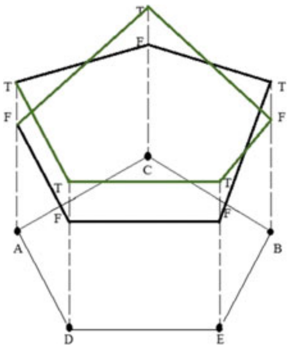


Fig. 7 Bundle diagram: liar cycle of five

We now extend the bundle diagrams. We depart from the liar cycle analogy, by saying that instead of truth values up the fibres, we have models. So, we now put on our classical, realist, model theorist hat. Our base space is now made up of formulas that are satisfied by models that are represented by stops up the fibres. It is paraconsistently (note! we shall return to this in the conclusion) possible to have mixtures of formulas whose models are locally consistent but globally inconsistent. In the 'model-extension' of the bundle diagram, the functions mapping models under formula (A) on to models under formula (B) and so on, fail to make a global assignment. There is a class of models that satisfies some formulas but if a particular subclass is satisfied, then some of the first formulas are no longer satisfied by those models, and the fibres are traversed more than once along their path.[19]

Changing the vocabulary to the more constructive, proof-theoretic version: instead of models, we can discuss possible inferences from a set of suppositions or hypotheses. If we cannot make a global assignment, then we 'need' (under our respect for the assumptions concerning the practice of mathematics, that mathematicians think of themselves as thinking consistently, classically or constructively) something like *C&P* to reconstruct the reasoning.

5 Bundles and Chunk and Permeate

Originally, the main purpose of introducing *C&P* and bundle diagrams was to formally depict mathematicians' reasoning when they are using inconsistent premises.

[19]We do not know if this might also be due to a reflexive iteration that causes what Dummett calls an 'indefinitely extensible concept'. This might correspond to the idea that we do not know if we are in an inconsistent cycle, so the edges might spiral up, but we have no way of knowing at any one point if we might then be brought down again. A bundle diagram where the edges spiral upwards indefinitely might represent something like a fractal where a new value is generated as a result of both the formula and the last value or last few values. This is all speculation that requires further investigation. We thank Jean-Paul van Bendegem for asking about spiralling edges.

In the classical mathematical proof case of the reconstruction, the *base space* is the premises and the conclusion. The stops on the fibre represent values for the variables of the theory. The values are either (i) the set of truth values that we can assign to the premises or other formulas in the language, this is all we need if our logic is only propositional or (ii) the (open) set of models satisfying the premises or conclusion, we need this for a first-order classical theory or (iii) the inferences that can most immediately be made (under some normal form and ordering of inferences) from the premises and the conclusion. We need this sort of stop up the fibres if we are reasoning in a first-order constructivist theory or proof theory.

A chunk can then be created by gathering the information in the base space below a *local* path that traverses *some* fibres each only once and traverses no fibre twice. Now we pay attention to the notion of a path being directed. This is not strictly necessary, and in some cases will not be appropriate; but it helps for the description here. Find a path that leads to the conclusion. The conclusion is in the last chunk, the *target* chunk. There will also be some premises or theorems in the target chunk. But the target chunk cannot include all of the premises. Information permeates, so there will be some overlap in information between the chunks. To keep things simple, try to minimize the number of chunks. In fact, it will often be possible to have only two chunks, depending on what information has to permeate to the other chunk. In bundle language: chunks consist in elements of the base space that are together consistent. So, as a first approximation, we individuate chunks as the base space below an edge that does not cross itself. Because some information permeates, there will be elements of the base space that find themselves in more than one chunk. That is what permeates.

Let us be more precise about what permeates. As we saw with the bundle representations, pairs of variables in the *base space* are locally consistent iff there is an edge between them. Following the bundle representation, we allow permeation and individuation of chunks in a 'back-and-forth' play between: on the one hand wanting to maximize chunks by taking the longest locally consistent assignment, and on the other hand, letting only consistent-with the-next-chunk-and-used-in-the-next-chunk information to permeate to the next chunk. So, now we are not maximising, but optimising between two considerations. A given maximal chunk might have to be made smaller, for reasons of permeation. There is some artistry here. This is a casualty, or strength, of our giving more specific guidelines than were hitherto available, while not wanting to make too formal and rigid the specifics of the bundle diagram. Situations that are amenable to more formal representation can be given a more effective recipe for individuating chunks and determining permeation.

Under our guidelines, we might break up our premises into sub-premises thus changing the base space. This corresponds to weakening axioms or splitting axioms into two. For example, in our liar cycle we could separate the special sentence from the quantified sentence (which would not include the special sentence, so the quantifier is bounded in an odd way!). Ignore this possibility, since it makes the notion of 'maximizing' or even 'optimizing' more complicated. It is because of such added complications that the method we propose here is not effective.

6 Generalizing Further: Bundled Chunk and Permeate to Reconstruct *Scientific* Reasoning

We generalize further. In mathematical reasoning especially today, we are fairly clear about what our premises are, where borrowed lemmas and borrowed theorems come from and what our concluding theorem is. In science, these matters are not always so clear. Moreover, we might be reasoning, not in the sense of deriving a theorem, albeit informally, but in the sense of reasoning about a concept, a phenomenon or a structure and so on, in order to deepen understanding rather than come to conclusions of deductions. Regardless of the difference in purpose, we might still be concerned about reasoning in an inconsistent context, where the ideas we bring to bear, in order to deepen our understanding, contradict each other under some representations in a formal language.

We shall work through an example very soon. Staying at the very general level for now: to deploy the *Bundle Chunk and Permeate* strategy (henceforth, *BC&P*) for reconstructing, or even guiding future reasoning, we start with deciding on the *base space* of contexts. For *BC&P*, *contexts* will be sets of premises, ideas, theorems, results, data from observations, or descriptions of phenomena which are jointly used when solving specific problems;[20] these sets will often coincide with the chunks (if already specified). These premises/ideas/theorems/results could be of two types: context–dependent, their interpretations and constraints are determined by the context in which such premises are being evaluated, or context-independent, their interpretations and constraints are given independently of which other premises are being evaluated (and, sometimes, the premises' value may be fixed and seem self-evident to scientists).[21]

We then have to ask a very fundamental question about valuation, in order to determine what we shall find going up a fibre. Valuations might be measurements, or they might even be qualitative, in the form of properties. An edge will connect values in adjacent fibres when the corresponding joint outcome is possible ([1], p.7). A global assignment will be indicated by a closed path traversing all the fibres only once. A contradiction could be –partially- pictured by assigning two *mutually incompatible* values to the same premise/idea/theorem/result. But that is not enough, to show the presence of global inconsistency, we need to show a 'twisted space' in topological language ([1], p. 1–2). As it has been described in [1], such a twist will help us to visualize the lack of a global assignment and the presence of mutually incompatible contexts.

We have identified at least four types of path over a global base space:

(1) An open path that does not cross all the fibers. This indicates nothing about the global base space.

(2) A closed path that crosses all fibers only once. This indicates a global assignment and with it, the possibility of consistent reasoning in the global theory.

[20] For related notions see [1], p.1.

[21] For related notions see [1], p.6.

(3) A closed path that crosses some or all fibres twice or more. This indicates that the valuations change as the reasoning along edges is carried out: this shows contextual reasoning, but not necessarily a contradiction. There is just a feed-back loop (recursive reasoning) that changes the valuations next time that you consider the base information. This is enough for identifying *local contextuality*.[22]

(4) A closed path that forces you to cross at least some of the fibers twice (or more). The segment carved out by the edges makes a twisted space: one where there is inconsistency in the global base space. To force the crossing twice there has to be a cross-over *between* fibers. This shows that the base space contains a contradiction. It shows global inconsistency.

In the following sections we present an application of *BC&P* for modelling inconsistent reasoning in empirical sciences. In order to do so, we shall first provide some definitions so as to make it easier to understand a particular case of inconsistent science and the particular application of *BC&P*. Then, in (Sect. 6) we shall introduce a case study from nuclear physics. And finally, we shall proceed to illustrate how *BC&P* could give a satisfying account of this particular case.

7 Some Preliminaries from Empirical Inconsistent Science

While paradoxes and internal inconsistencies in the more formal sciences are well-documented in the literature and have called the attention of many paraconsistent logicians; inconsistencies from empirical sciences (in particular, inconsistencies between theories or models)[23] have not enjoyed as much attention. Some exceptions are mentioned in this paper. We think that the presence of some contradictions in scientific reasoning is not a minor issue; so, if a case of inconsistent (non-trivial) science is spotted, it seems necessary to offer an explanation about how inconsistent information can be combined and not become trivial in the empirical sciences. We believe *BC&P* can help us to achieve such an explanation or reconstruction of the reasoning.

7.1 Different Groups of Propositions

In the empirical sciences, the different disciplines and research domains are never completely independent of each other ([27]: 53). As a matter of fact, more often than we expect, in actual scientific practice different theories (from different disciplines) and different models (from different theories) are often combined for solving specific

[22] Characterized in [1] as: "there is a local assignment which is in the support, but which cannot be extended to a global assignment which is compatible with the support." (2015, p.6).

[23] Schummer ([48], pp. 64–5) argues convincingly that, especially when considering problems in chemistry, we use the term 'model' rather than 'theory', since this better reflects the practice of chemists when reasoning about phenomena in chemistry. Of course, here we do not mean model in the model theory sense of the term.

problems. Some problems are complex enough that they cannot be clearly solved by one theory or model alone.[24] We shall focus mainly on inconsistencies that involve two or more different—original—groups of propositions.

That being said, a natural question emerges: how can scientists individuate groups of propositions as 'distinct theories' or 'distinct models'? This is not a trivial question. When we individuate objects we give the necessary and sufficient conditions under which we are able to tell when two objects are different from each other, and when, what we thought were two distinct objects, turn out to be the same one. Famous examples include the discovery that the evening star is the morning star or that 'jade' is really two different chemical compounds: now called 'jade' and 'jadeite'.

Nonetheless, individuating scientific theories along the sorts of standard lines we would use to individuate mathematical theories is disingenuous towards scientific practice.[25] When talking about scientific theories the challenge is double: it is difficult to say when a set of objects, substances or ideas is different from another set of objects, substances or ideas, and it is also difficult to specify when two set of objects, substances or ideas are part of the same scientific theory [38]. As a matter of fact, the history of philosophical and scientific debates has shown that sometimes

> we can't agree on which set of 'things' constitute 'Newtonian cosmology', 'classical electrodynamics', and the rest. We see in (...) particular examples of disagreements about whether some theoretical constituent (equation/model/proposition) should or shouldn't be considered 'a part of the theory'. ([40], 2892).

Such disagreements are hostage to the abstract activity of theory individuation. When philosophers or scientists ask themselves if a particular scheme is 'really the theory X', there is often miscommunication ([39] chap. 2, [40]). However, if they ask themselves which are the theoretical constituents that are sufficient to solve a particular problem given certain constraints, agreement is reached more easily.

Taking that into account, we take a naturalist stance, and observe that often in actual scientific practice, scientific theories are not individuated abstractly, but in terms of specific problem solving goals. Here we shall claim that a theory (or model) will be successfully *individuated according to a particular problem*, if the set of propositions that constitute such a theory (or model) entails a solution to the problem, or a statement of the problem as well as a neat, or systematic understanding of it.[26] Once a theory is individuated, it could be studied by analysing it *globally*, this is, through the revision of the properties that the whole theory possesses; or it could be studied through the analysis of some of the properties that only some of its subsets possess.[27]

[24]For examples of this see Elsamahi [36, 37].

[25]If we were to be normative, or even prescriptive, about science we could disregard scientific practice and force individuation of scientific theories in order to avoid inconsistency within a scientific 'theory'. We do not propose to do this, since, as we shall see, this would be quite unnatural to the practice, and of rather limited interest.

[26]See ([27]) for related notions of *problem solving*.

[27]Even though almost any scientific theory could be fragmented in infinite ways, here we shall focus only on such subdivisions that are compatible with the way in which scientists use their theories in

A theory could be separated into *meaningful subsets* if and only if the elements contained in such subsets are considered to be sufficient for solving interesting problems in the discipline to which they belong—if they are too minimal for solving problems, we will not consider the separation to be a candidate for being a chunk. The study of the properties that meaningful subsets of the original theory possess is what we understand as *local* analyses. The properties that are present in meaningful subsets of a theory, are not always present in the theory as a whole. For instance, a theory could be locally consistent, i.e. could have consistent subsets, without necessarily being *globally* consistent.

A particular theory (or model) *A* will be distinct from another theory (or model) *B* according to a particular problem if and only if the solution of the problem that is entailed by the theory (or model) *B* cannot be achieved without the theory (or model) *A*.[28] Two distinct theories (or models) could be *satisfactorily combined* if and only if they are distinct theories and if their combination allows for larger explanatory or predictive power, solving other problems or contribute to greater understanding, than the one that each theory alone possesses.

In sum, we consider that the relationship between a particular group of propositions, their theoretical context and the solution to a specific problem are all necessary for individuating theories (or particular models).

This way of individuating theories is not only loyal to the practice, but it will be useful for separating groups of propositions when applying *BC&P* to particular cases.

7.2 Global and Local (In)Consistency

As we claimed in the introduction, scientific inconsistent theories have often been analysed at two scales of analysis: global and local. What interests us here are cases where the global theory, or model is inconsistent, but sub-theories or set of models are consistent.

While we take a 'naturalist stance' in the sense of respecting the scientific practice and allowing it to guide our analysis on the relations between different theories and between different models, we should note that respecting the practice comes at a price: many of the scientific theories that are built and individuated under problem solving considerations are, at some point in their development, inconsistent. For example, Bohr's theory of the atom was initially designed for explaining why hydrogen emits and absorbs light at certain specific frequencies and, since the beginning, the theory

their standard practice. Henceforth, we shall refer to this way of choosing chunks as separating it into *meaningful subsets*.

[28] In what follows, we shall assume that scientific theories are often individuated following specific problem solving considerations, and that this individuation is often in terms of objects, sets of phenomena, sets of forces acting together, or classes of axiomatic theories, among others.

succeeded at its main goal. However, despites this success, the early versions of the theory were inconsistent ([11, 17]).

As a matter of fact, the list of theories that (allegedly) have been inconsistent is long and diverse, some examples of inconsistency in science are Bohr's theory of the atom ([11, 17]), the Early Calculus ([5], [25, 59]), Classical Electrodynamics ([19, 20]), Prout's hypothesis [43], the models of the atomic nucleus [37], among others.

Although all those case studies aim at illustrating inconsistent scientific theories, some logicians and philosophers of science have pointed out that the inconsistencies that have been portrayed by these cases are not really homogenous ([15, 27, 33, 43]). As a matter of fact, "if we distinguish between observation and theory (what cannot be observed), then three different types of contradiction are particularly noteworthy for our purposes: between theory and observation, between theory and theory, and internal to a theory itself." ([43], 144).

These differences play a crucial role in the philosophical analyses of inconsistencies in the empirical sciences. Nevertheless, here we are analysing inconsistency in science as a logical concept.[29] We shall gloss over the differences by focusing on sets of sentences or formulas. Thus, the sets of sentences might be about observations and theory, might belong to 'different' theories or might all belong to a theory. Since we are interested in representing the inconsistencies using the bundle diagrams, the sets of sentences are our variables. They make up the base space.

We then focus on the distinction between local and global. These could be characterised as follows. Given a specific problem X, and two different groups of propositions,[30] a and b,[31] that are put together to provide a solution for X:

- a is locally consistent if and only if a does not contain nor entail a contradiction.
- b is locally consistent if and only if b does not contain nor entail a contradiction.
- The union of a and b is locally consistent if and only if the union does not contain or entail a contradiction.[32]

While in the sciences (formal and empirical) it is often expected that the union of two locally consistent sets of information is still consistent, because true and about the world, this is not always the case. However, we shall show that it is not as dangerous as has been traditionally thought ([24, 42]). In what follows, we shall provide a case study from nuclear physics where we combine two theories or models, that each is internally consistent, but their union is inconsistent. Moreover, the union is needed for having a more complete understanding and to solve some problems in science.

[29] For our bundle diagrams these differences would be drawn out by our choice of variables: be they observations, theories or ideas within a theory.

[30] The propositions could be empirical assumptions, observational reports, laws, theorems, axioms, etc.

[31] Here a and b could be either distinct theories or distinct meaningful subsets of the same theory.

[32] Of course, the union of two locally consistent sets of information need not be consistent with each other.

8 A Scientific Example of Bundled Chunk and Permeate

The first case studies that were modelled by using *C&P* were cases of internal inconsistency, but recently, an example of a different kind has been provided as a candidate for using *BC&P*. The example is the combination of two mutually inconsistent climate models that allow for accurate predictions regarding temperature, pressure, humidity and other meteorological quantities [7, 9]. We consider it important to emphasise the fact that scientists very often make use of mutually contradictory bodies of knowledge in order to solve problems in their discipline, here we shall introduce a similar case study from nuclear physics. We chose to present this particular case to reach three main goals: to introduce a new case of inconsistency toleration in empirical sciences (Sect. 8.1), to illustrate an application of *BC&P* (Sect. 8.2), and also, to draw some philosophical conclusions about inconsistency toleration and the unification of science (Sects. 9 and 10).

8.1 The Case Study

In a nutshell, the case study goes as follows: the Liquid Drop Model and the Shell Model contain incompatible basic principles regarding the structure of the nucleus of an atom; it is only when nuclear physicists combine some of the predictions of both models that they gain accuracy in their predictions and measurements of binding energies for all the chemical elements of the periodic table and in their predictions and explanations of other nuclear processes such as fission. This case study illustrates a scenario in which each model can accurately predict only a segment of the elements in the periodic table and only part of a general phenomenon, but in which combining the predictions of both models provides successful descriptions and predictions of more general phenomena.

First, the nucleus of an atom is the small region in which 99.9% of the total mass of the atom is located. The nucleus consists in protons and neutrons that are bound together. The protons are responsible for the positive charge of the atom. The behaviour of the nucleus is explained by appealing to two different forces: the strong nuclear force and the weak nuclear force. The strong nuclear force is what binds nucleons (protons and neutrons) into atomic nuclei, while the weak force is responsible for the decay of neutrons to protons. Any atomic nucleus (of any chemical element) will exhibit binding between protons and neutrons and decay of neutrons and protons.

The binding energy of a nucleus is what in large part determines the stability of the nucleus. Ideally, binding energies are necessary for understanding and determining under which conditions a nucleon can change to another (from neutron to proton, for instance) or escape from the nucleus. Considering that binding energies are necessary for predicting and describing different aspects of the nuclear structure (for instance, correlations present in the nuclear ground state [16]) physicists have tried to come up with a homogeneous theoretical framework to calculate this type of energy.

Our current nuclear physics provides us with models of features that allow us to, at least, describe, predict and measure this type of behaviour of atomic nuclei. Such models have been developed by different research programmes that have a main goal in common, namely: to provide some insight into the structure and dynamics of atomic nuclei. Today, there are 31 different successful and internally consistent nuclear models that offer some insight into the nucleus of the atom (Cf. [14, 37]). These models are often classified into three main groups: microscopic models (focused on nucleon-nucleon interactions), collective models (focused on bulk properties of the nucleus as a whole) and mixed models (which are somewhere in between the two previous ones).[33] However, as yet, there is no consistent or coherent global account of the structure of the nuclei that allows us to explain, predict and measure all of the nuclear behaviours.

The diversity of models itself is not problematic; especially if "each model has its particular successes, and together they are sometimes taken as complementary insofar as each contributes to an overall explanation of the experimental data" ([37]: 179). However, the case study that we are presenting here, illustrates how the basic assumptions required by one model contradict those required by another model ([14, 32, 37]), more important, none of these conflicting assumptions seems to be idle, and they all are, allegedly, strongly linked to success in particular applications of each model [37]. Let us press this point further by describing two such mutually incompatible nuclear models.

The first of these two models is the Liquid Drop Model (*LDM*). It is one of the most successful nuclear models. The LDM was formulated more than 80 years ago under the assumption that the nucleus of an atom exhibits classical behaviour (protons and neutrons strongly interact with an internal repulsive force proportional to the number of nucleons). The model was based "upon the experimentally established dependence of total binding energy of a nucleus upon the number of nucleons. As expected for a liquid, the nuclei proved to be almost incompressible and their total binding energy included a negative term, proportional to the volume of a nucleus, and a positive term, proportional to its surface" ([50]: 219). Since the beginning, the *LDM* could predict and describe a series of nuclear properties, such as the growth of the nuclear charge, the instability related to Coulombic forces, the evaporation of nucleons after heating, the nucleus' change of shape, and the phenomenon of spontaneous fission, among others ([14, 37, 50]).

However, despite its success, the *LDM* fails to describe the way in which the nucleus often displays distinctive energy levels forming shells and subshells (the so-called shell effects). It also fails to give a full account for the ground-state properties of nuclei ([49, 50]). Additionally, "the quantitative description of the nuclear force that emerges from nucleon-nucleon reaction studies is incompatible with what is known about nuclei" ([37], 178). Finally, while the *LDM* can be used to predict and describe binding energies of nuclei of any element of the periodic table, in the

[33]This classification was first developed in ([22]), and later in ([14]).

corresponding experiments, some nuclei show systematic deviations with respect to the *LDM* predictions. Experimentation has shown that some nuclei are bound more tightly together than predicted by the *LDM* depending on the number of nucleons that they possess. To explain this phenomenon, scientists refer to the so-called 'magic numbers'. The phenomenon can be detected in the nuclei of atoms of, at least, Helium (*He*), Oxygen (*O*), Calcium (*Ca*), Nickel (*Ni*) and Lead (*Pb*).

Nonetheless, the partial failure of the *LDM* does not mean that scientists are left empty handed. When dealing with the phenomena that the *LDM* cannot describe, physicists often rely on other models, one of the most important is the Shell Model (*SM*). This nuclear model was formulated more than 70 years ago and aims at describing and predicting, among other nuclear properties, the shell effects of the nuclei. In this model, a shell represents the energy level in which particles of the same energy exist, and so, the elementary particles are located in different shells of the nucleus. According to the *SM* the nucleus itself exhibits quantum-mechanical behaviour. "The basic assumption in the nuclear shell model is that, to first order, each nucleon (proton or neutron) is moving in an independent way in an average field" ([16]: 58); that is, for this model "nucleons are assumed to be point particles free to orbit within the nucleus, due to the net attractive force that acts between them and produces a net potential well drawing all the nucleons toward the centre rather than toward other nucleons." ([37]: 185). One of the most important virtues of the *SM* is that it accounts for the magic numbers phenomenon, among other important experimental data.

Considering the diversity of models and the obvious conflicts between them, nuclear physicists have untiringly attempted to combine both microscopic and collective models in order to provide a unified framework of the behaviour of the nucleus ([14]). Common manoeuvres have been related to the combination of elements from the *LDM* with elements from the *SM* (Cf. Groote, Hilf and Takahasi 1976; [14, 16, 50]), however, the success of any of the attempts is still unclear.

A large number of nuclear physicists agree that "material systems such as nuclei are too complex and contain too many constituents to be handled precisely with formal "bottom-up" theories, but they are too small and idiosyncratic to be handled with rigorous statistical methods that normally require large numbers to justify stochastic assumptions" ([14], 57), and in that sense, even if endorsing unificationist commitments, physicists take for granted that nowadays there are some scientific problems whose solution requires the use of more than one nuclear model. For instance, the calculation of binding energies of all elements of the periodic table; which, for accuracy, requires the use of the *LDM* for almost all the elements, and to use the *SM* for those nuclei with magic numbers. An important remark: to provide accurate predictions concerning binding energies is not an idle task for nuclear physicists, especially in light of the privileged role that such energies play when describing, calculating and explaining nuclear processes such as fission.

So, if they want to address the domain of binding energies of all the chemical elements, physicists have to agree that at present there is no single direct way to calculate them; instead, we have to use two mutually contradictory models, each one of them accurately predicting only a segment of a general phenomenon. The contradiction involved is even more troubling when we consider that both models contradict each other about the structure of the nucleus, and that such characterisations of the nucleus are, allegedly, what is largely responsible for the success of each model in particular applications ([37], Chap. 5).

Nuclear physicists use both models, *LDM* and *SM*, to calculate the binding energies for all elements of the periodic table; later on, they use such results for predicting nuclear reactions such as fission. They calculate binding energies of the nuclei with magic numbers using the *SM* and (for simplicity) use the *LDM* for the rest.

If scientists want to reason classically or constructively, and avoid triviality when solving these problems, they either have to get rid of some basic assumptions of specific models (by deciding that they are in fact idle, for instance), or they have to find a way to connect the consequences of both models without allowing explosive reasoning. Insofar as what has been said here is correct, this example from nuclear physics is a good candidate for being modelled by *BC&P*.

8.2 Nuclear Physics and BC&P

For simplicity, here we shall only illustrate the case of nuclei of Helium-4 (*He*4). First, the individuation according to a particular problem goes as follows: the *problématique* that requires explanation is the behaviour of the atomic nucleus, in particular, the phenomenon of fission of nuclei of *He*4. The theoretical constituents that are sufficient for solving that problem are the *LDM* and the *SM*. For, atoms of Helium-4, we also include information about how *He*4 is one of the nuclei with magic numbers, as well as the fact that the nucleus of *He*4 is identical to an alpha particle.

The basic assumptions of the *Liquid Drop Model* we need are, at least, the following:

(*D*1) The nucleus behaves as a classical fluid consisting in protons and neutrons that strongly interact with an internal repulsive force proportional to the number of protons.

(*D*2) Nucleons move randomly and bump into each other frequently.

(*D*3) The nucleus itself exhibits classical behaviour. (Cf. [13, 37])

(*Dn*) The *semi-empirical mass formula*[34]:

$$E_b(MeV) = a_v A - a_S A^{\frac{2}{3}} - a_C \frac{Z^2}{A^{\frac{1}{3}}} - a_A \frac{(A-2Z)^2}{A} + \delta(A, Z)$$

(*Dc*) The *LDM*-predictions regarding fission of *He*4 nuclei.

The basic assumptions we need from the *Shell Model* are, at least, the following:

(*S*1) The nucleus exhibits a quantum-mechanical behavior.

(*S*2) The atomic nucleus is a quantum n-body system.

[34] The formula is based on the *DLM* and is used to predict binding energies of nuclei. It is also called "Weizsäcker's formula".

(*S*3) The nucleus is not a relativistic object and its equation of motion (the system wave function) is the Schrödinger equation.

(*S*4) The nucleons interact only via a two-body interaction which is, in effect, a practical consequence of the exclusion principle.

(*S*5): The nucleons are assumed to be point particles free to orbit within the nucleus, due to the net attractive force that acts between them and produces a net potential well drawing all the nucleons toward the centre rather than toward other nucleons. (Cf. [37]).

(*Sc*): The *SM*-prediction of the *He*4 binding energy,

We might think that we should make two chunks, one for each model with the common information permeating from one chunk to the other, but in standard explanations in nuclear physics, we more naturally find four source chunks:

- *Einput*: contains the empirical data about *He*4 nuclei, including that it has a magic number.
- *LDM*: contains the assumptions of the *LDM*, *D*1,*D*2 and *D*3 and *Dn*. This chunk will *grow*, as we let in data contained in *Einput* and obtain as a result the *LDM*-predictions regarding fission of He4 nuclei (*Dc*).
- *SM*: contains the assumptions of the *SM*, *S*1,*S*2,*S*3,*S*4 and *S*5. This chunk will *grow*, as we let in data contained in *Einput* and obtain as a result the *SM*-predictions if the *He*4 binding energy (*Sc*).
- *Exp*: contains the experimental reports on binding energies of He4. Of *Einput* and *Exp* chunk, one is (locally) true whenever the other is (locally) true, and they are always assumed to be so.

Our *BC&P* reconstruction recognises also one target chunk, the solution to a problem:

- *Eoutput*: contains the empirically adequate predictions concerning binding energies and fission of He4 nuclei.

That considered, the base space of our *BC&P* includes four source chunks (*Einput*, *LDM*, *SM*, *Exp*) and one target chunk (*Eoutput*). In addition, along each of the fibres are four stops, which represent the possible valuations considering the two main goals: first, to calculate the binding energy of *He*4, and second, to calculate fission for the *He*4 nucleus. The first two stops indicate if the statement that is evaluated is considered to be true (*Tb*) or to be false (*Fb*) when calculating the binding energy, the second pair of stops, indicate if the statement is assumed to true (*Tf*) or to be false (*Ff*) when predicting fission.[35]
Now, when pursuing the target chunk, it is necessary to first determine the binding energy of *He*4 nucleus. In order to do so, we first assumed that the sentences contained in *LDM* are false, and then assume that the data from *SM* is true. We start to construct our bundle diagram to represent this. See Fig. 8.

[35]Note that *Tb* and *Fb* are mutually exclusive, and the same goes for *Tf* and *Ff*. Nonetheless, the following pairs are mutually compatible: *Tb* and *Tf*, *Tb* and *Ff*, and *Tf* and *Fb*.

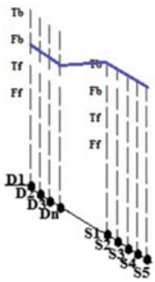


Fig. 8 Bundle diagram: basic assumptions of LDM and SM

That this is compatible with *Einput* being true –especially considering that *Einput* includes the concept of *He*4 having a nucleus with a magic number. See Fig. 9.

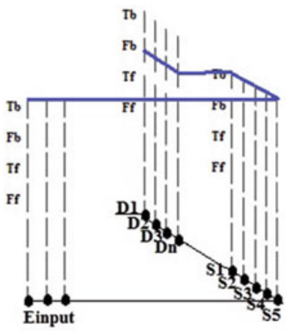


Fig. 9 Bundle diagram: adding empirical input

Due to the falsity of *LDM* , and the assumption of sentences in SM being true, it is according to our scientific reasoning, allowed to combine *SM* with *Einput* to obtain *Sc* –which is taken as true. See Fig. 10.

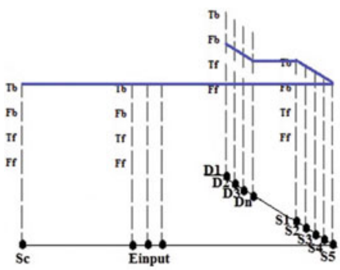


Fig. 10 First result: Binding energy

As Sc coincides with what is contained in *Exp*, and so, what is expressed by *Sc* we are then is allowed to move to the target chunk, *Eoutput*. See Fig. 11.

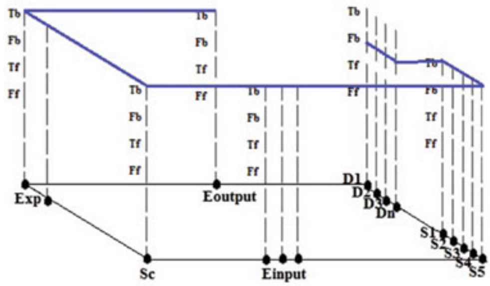


Fig. 11 Contrasting with the experimental results

Once the accurate prediction of the *He*4 binding energy is available, the next step is to explain and predict fission for *He*4 nuclei, and for that, physicists will use the *LDM*. Thus, the sentences in the *LDM* chunk are taken as true. See Fig. 12.[36]

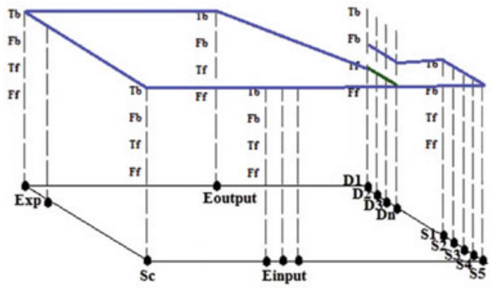


Fig. 12 Contextuality and the beginning of the second stage

Figure 12 is a representation of a base space that includes: propositions of each model and predictions of each model regarding nuclear reactions of *He*4 nuclei, as well as a general description of the phenomena of fission and binding energies (regarding *He*4 nuclei). We need the whole base space to predict fission for *He*4 nuclei. There is an edge between variables when they can be used together in order to enable measurements of binding energies. The fact that there is no closed path traversing the fibres only once connecting particular base space points, such as

[36]Note that we have changed the color to indicate that we have moved to the next step in the calculations involving nuclear fission for *He*4 nuclei. As it is in scientific reasoning, the edges have direction, *Sc* has to be moved into *Eoutput* before it is possible to make any prediction regarding nuclear fission. This is new (to the bundle diagram construction) but it is inherent to standard scientific reasoning.

Eoutput and $D1$, shows the *logical contextuality* of the model. It shows that there is no global assignment that allows for recuperating the phenomena of binding energies and nuclear fission as a whole: for instance, when $S1$-$S5$, Sc and Exp are true of the phenomena, $D1$ to $D3$ and Dn cannot be true. However, at the end, *LDM*-assumptions and *SM*-assumptions are both necessary for giving an account of the general phenomena of fission for *He*4 nuclei.

Finally, once again, what is in *Einput* is taken as true. And because *LDM* is true, *SM* is taken as false.[37] *LDM* is combined with what is in *Einput*, and it is possible to obtain Dc (predictions about fission). Due to the compatibility between *Exp* and *Dc*, *Dc* is allowed to flow to the target chunk, *Eoutput*. Now, in *Eoutput* nuclear physicists have both the predictions of binding energies for *He*4 nuclei as well as the ones for nuclear fission for such atoms.[38] See Fig. 13.

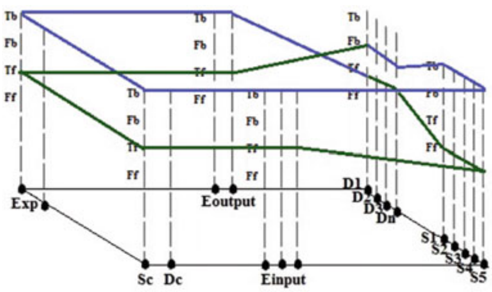


Fig. 13 Bundle diagram: binding energy + fission

What the bundle diagram has shown is revealing in more than one sense. First of all, it allows us to see that even though the target chunk, *Eoutput*, can be somehow constructed, it only happens in a very conceptual and abstract sense, where we assume that some information used for the construction is false ($D1$ to *Dn*) or ($S1$ to $S5$). The twist between *Dn* and $S1$ shows logical contextuality and the presence of only logical contextuality.

In addition, the diagram can also suggest which chunk is fully compatible with which other chunk (for instance, *Einput* is fully compatible with *SM* and with *LDM*, but it is also fully compatible with *Exp* and with *Eoutput*) and also suggests in which cases information ought to be filtered (for instance the fact that *SM* and *LDM* are mutually contradictory and they both still feed the target chunk, clearly suggests that only limited information should be allowed to move from such chunks to the target one).

[37] *Sc* is kept as true because it is compatible with the empirical assumptions (*Einput*), the experimental reports (*Exp*), and the *LDM*-predictions for fission, and also because it is now part of the target chunk.

[38] *DLM*-assumptions are next taken as false in cases in which the next move is to predict other properties of nuclei, such as spin and parity of nuclei ground states.

Finally, the use of *BC&P* diagrams can also help us to see how each of the nuclear models is locally consistent. When looking at the areas in the diagram that correspond only to each model, we can identify a local assignment that corresponds to a closed path traversing all the fibres over that part of the base space exactly once; and the same happens for the shell model. However, as should be clear to the reader, here, the local consistency comes at the price of the impossibility of using only one model for predicting both binding energies and fission for atoms of a certain type (those with magic numbers).

9 Problems with Global Inconsistency in Science

There are two related problems. One is that it has long been presupposed, especially in the more Western scientific traditions,[39] that *good* reasoning should be consistent ([15, 24, 42]). The other problem is that it is often assumed that, in the long run, our best science can be, and *should* be unified into one body of knowledge; where unification presupposes consistency. One can find several research projects that are in line with this particular pretension, for instance, the current project for unifying the four fundamental forces. Call the first 'the meta-logical problem' and the second 'the unification problem'. The second presupposes the first.[40]

Put another way, the first problem is that if we were to meet a contradiction in our science, then reasoning would be impossible. We inherited this meta-logical idea at least from Aristotle [44] if not from before. We call this presupposition '*meta*-logical' because it concerns the limitations of logic. We call it a *presupposition* because there are perfectly rigorous formal systems of reasoning that include contradictions as features, so it is unnecessary.

Why are contradictions so detrimental to reasoning in science? Rehearsing what we learned before in Sect. 1, if we are classical or constructive reasoners, then we endorse *ex contradictione quodlibet* reasoning as *valid*. Assuming classical or constructive reasoning in our science, in the face of a contradiction, we have explosion. Explosion in a theory means that every sentence written in the language of the theory, or every formula written in the language of the theory is true, or is derivable. So if the langage contains some form of negation or denial, then a sentence and its opposite are both true, or both derivable. This is what we call trivialism. The problem with trivialism is that it is undiscerning. There is no error and so no correction possible. Anything goes (within the constraints of the language).

What does this mean for science? In a science this might well also include observation sentences. So, we might observe that the temperature indicated on the thermometer is roughly zero, but also that it is one hundred degrees, or roughly seventeen

[39] Arguably, in more Eastern traditions of 'science', contradictions are tolerated, (Garfield: Engaging Buddhism). Such Eastern 'science' might not be recognised to be science at all *a priori* because reasoning with contradictions is a priori impossible.

[40] Of course, it does not meta-logically have to, but it happens to.

degrees, all at the same time and in the same situation. This makes nonsense of our 'science'. Of course, note that this is all qualified by the antecedent of the conditional of the second sentence of this paragraph: that we are classical or constructivist reasoners.

Turning to the second problem: it is that we have an ideal towards which we strive as scientists, and this is to unify science. What does this mean? Philosophically, it is presumed[41] that the real world is a 'unified' place, and this intimately includes the presupposition that 'reality' is not contradictory.[42] To our delight we have also found that our scientific theories 'work' and that we have tangible progress in science. This falls in line with the 'cumulative retention' tradition in philosophy. That is, science serves us to predict, explain and control our natural environment. So, it is in this sense that we have 'success' in science.

> [V]irtually all models of scientific progress and rationality (with the exception of certain inductive logics which are otherwise flawed) have insisted on wholesale retention of content or success in every progressive-theory transition. According to some well-known models, earlier theories are required to be contained in or limiting cases of, later theories; while in others, the empirical content or confirmed consequences of earlier theories are required to be subsets of the content or consequence classes of the new theories. [28]

Under this ideal conception, our individual scientific theories represent parts of the unified (consistent) reality. It then follows that insofar as our theories reflect reality, they *should* be consistent, not only within themselves but also with each other. The unity of science consists in the global scientific project of making one consistent theory that predicts and explains the whole of our natural world. So, rather than separating the liquid drop model from the shell model of the nucleus of an atom, we should be able to seamlessly reason from one to the other without meeting contradictions. Of course, to arrange for this seamless reasoning, we would have to alter the theories. Under a unified science, the distinctions between theories would then be a matter of history and convenience; both conceptual and institutional. A unified theory of the whole of scientific reality would consist in one set of laws from which we would derive natural phenomena given some initial data.

In the concluding section, we shall question the presumptions made concerning the unity of science, but for now, we recognise it as intrinsic to some of the practice of science. An exception is chemistry [48]. Recognising the ideal of science, that it should eventually be unified, we can see the problem if we find contradictions in science. If there are contradictions within and between our theories, then this, by definition, impedes unification. To think of contradiction as such an impediment depends on the presumption that led to the first problem.

[41] We shall show exactly why this is a presumption, and on what it rests.

[42] For a short discussion of this issue where the possibility of a some-places contradictory real physical world, see [21].

10 Philosophical Conclusions

We have been trying to 'make sense of' mathematician's and scientist's reasoning with inconsistent premises or statements from theories or models. 'Making sense', here, means that we want to make a rational reconstruction under the pressure of scepticism that such reasoning is illegitimate. Such pressure arises when a proof is relatively informal, and uses information from different mathematical or scientific theories that we know are incompatible with each other, where 'incompatible' means that we know or suspect that if the two theories were written in the same language, then it would be possible to derive a contradiction from the two theories.

Do we see such reasoning in mathematical or scientific practice? Yes. We have done so for a very long time. The example we are most familiar with are those of the early calculus and of Lobachevsky solving a problem in Euclidean geometry (about the space under an indefinite integral) by appealing to his hyperbolic geometry. The latter case is a rather simple one for *C&P* to work with, since there is no information about parallel lines that is used in the proof. This corresponds to the cases we referred to in the introduction where premises come from inconsistent theories, but are not themselves inconsistent with each other. More important and blatant uses of inconsistent premises are reconstructed in [10]. We have shown similar reasoning in physics where we mix the liquid drop model with the shell model. Increasing numbers of PhDs in mathematics are written using informal proofs that borrow from different areas of mathematics; similarly for science. Since these are original, and mark new territory in mathematics and science, they are exactly the sorts of proofs, or types of reasoning, we should be cautious, and sceptical, about. In physics, chemistry and biology we only have incomplete theories. The great unification of the sciences is not on the horizon. Even the methodologies are sometimes in direct competition [48].

For example, both the Andréka-Németi group's [2] and the Krause and Arenhart [26] approaches to physics urge us to develop a logical reduction/explanation of physics. They disagree on what counts as a logic, and what counts as a reduction. The first group prefer a first-order logic without the notion of forces, and where 'causation' is simply expressed in terms of before and after on the trajectory of a body, or on a spatio-temporal relationship between two bodies. Any vestige of causation that they have is metaphysically bare.

In contrast, the Krause–Arenhart approach uses a higher-order language with proper classes and forces that are causal. The 'reduction' of the Andréka-Németi group is thought of in terms of several formal theories and the limitative relations that bear between them at the meta-level. The 'reduction' for the Krause–Arenhart approach follows Suppes, to have one set of axioms, so one logical/ mathematical theory. Thus, even here, where we have a highly mathematical, nay logical, approach to problems in physics, there is no promise of unification in a traditional sense of one theory. The fragmentation of science is in evidence. "*Quant à? l'unité de la science, si ardemment projetée jusqu'au début du xxe siècle, elle est finalement restée pure* *pétition de principe devant la spécialisation croissante des domaines scientifiques*"[43] ([29], 13).

Because of the fragmentation of the sciences on the one hand, and the need to use ideas from incompatible areas of the science to give fuller understanding and explanations, and to make better predictions and control on the other hand; it is pressing to reconstruct the reasoning, in order to show its coherence in the presence of global inconsistency. For the reconstruction we use an enhanced version of *C&P*—bundle informed chunk and permeate: *BC&P*. The main motivation for using *C&P* over a paraconsistent formal representation of the reasoning is to preserve the meta-logical intuition that scientists tend to share, that they reason either classically or constructively, and even if they are not able to articulate these meta-logical intuitions in these words, they would all find reasoning through a contradiction in science to be problematic. Abramsky (in private conversation) uses the bundel diagrams to counsel us to reason short of inconsistency, we should not reason using the inconsistency. What we add to *C&P* as it has been developed, is the bundle diagrams, as a guide to individuating chunks and selecting what information permeates from one chunk to the next.

In our particular example, we used the two models of the nucleus of an atom, the liquid drop model and the shell model. In the presence of measurements, we find that the models contradict each other. Nevertheless, both are needed to explain the phenomenon of binding energies. We thus extended the application of the bundle diagram to include not just arguments, but more broadly the relationship between models when they are both used in an explanation.

This was one example. We could extend the technique further. To do the *BC&P* reconstruction for non-model theoretic proofs, ones that more closely resemble proof theory, we would need to change the bundle diagram, so the 'values' are now, say, mediate inferences from suppositions, where we make the formulas unique *via* a combination of normal-form of language and by imposing some ordering on formulas.

The bundle diagrams represent a situation where we have global inconsistency. The same can be done, without diagrams using sheaf theory and cohomology theory [1]. Both tell us when it is 'safe' (i.e. consistent) to extend our reasoning, and when it is that we overstep the bounds of consistency. *C&P* helps us to stay just within the bounds: we can be systematically careful about what formulas, axioms, assumptions, measurement statements we can locally consider together, and which we cannot. So the bundle-extension of *C&P* can handle quite a lot of cases.

What cannot be handled? If we are using formulas, theorems, results from different theories or measurements where inter-translation is not obvious, it is not clear that we could come up with a bundle diagram, and it might be more work than it is worth. That is, it might be just as difficult to do this, as it is to generate some other meta-proof of local consistency. The limitations to such extensions concern deciding

[43] As for the unity of science, so adently pursued up to the beginning of the twentieth century, what has remained is nothing but a guiding principle, unatainable under the increasing fragmentation of the domains of science.

what the *variables* are, what count as *valuations* for the variables (since they might not be common) and in cases where the valuations are different up each fibre, what is to count as an edge, since the semantics is quite different, it is not clear how to make a translation to then determine if two elements of the base space are pairwise consistent.

For example, say, one premise comes from model theory, and another from proof theory, then the valuations for the model theory formula will be various models (note also that we might not be able to order them up a fibre, and this is another limitation), and up the proof theory 'variable's' fibre, we might have immediate and mediate inferences. For the purpose of drawing edges between fibres, we might need to have some sort of translation, and this might not be obvious or desirable in all cases.

Another criticism of this approach is that it might not remain loyal to the intended reasoning of the mathematician who came up with the proof in the first place, or of the scientist who came up with the mixed explanation. This is quite correct. All that *BC&P* promises is that it is a means of staying loyal to the idea that is wide-spread in the mathematical and scientific communities that local reasoning is consistent, and usually classical or constructive.

There is a more interesting and thorny issue that we are touching on. It is that while at the object-level, we are being careful to 'stay locally consistent' at the meta-level, we must be reasoning paraconsistently in the very limited sense that we are reasoning about reasoning consistently within globally inconsistent theories, or models. A bundle diagram which has no path traversing each fibre only once *represents* reasoning inconsistently. So we are looking at a diagrammatic representation of inconsistency and reasoning *about* inconsistency, and this might be thought of as 'reasoning paraconsistently' without reasoning using a particular paraconsistent logic.

Some paraconsistent logicians claim that mathematicians reason paraconsistently, unbeknownst to them, in exactly this way. The claims of the developers of *C&P* are a bit ambivalent about the relationship between the *C&P* strategy and paraconsistency. Brown recognises that we *could* use a paraconsistent logic within a chunk, in principle, although this was not his original intention. Also, we are reasoning paraconsistently at the meta-level in the thin sense that we recognise the presence of inconsistency, and want to avoid bringing about explosion. What the very possibility of *BC&P* reconstruction shows us is that what we immediately fear is explosion, and only mediately, indirectly, inconsistency. This is one of the lessons of paraconsistency. Moreover, we are at pains, at the meta-level, to make very clear the distinction between explosion and inconsistency through the bundle diagrams. If we can represent the inconsistency, and avoid it, by reasoning short of it, then we reason paraconsistently in the limited, thin, sense of exercising damage control over the inconsistency. The details of the reasoning could be captured using a formal paraconsistent logic, but this is unnecessary.

What is interesting is to draw the lesson that it is crossing the *ex contradictione quodlibet* boundary into triviality or detonating explosion that is otiose in the present practice, not the lingering background possibility—although this is enough to already upset more sensitive souls. In the practice of mathematics and science today: having a bomb and a detonator is fine, using it is not. So, what we have done with *BC&P* is give a means of vindicating inconsistency toleration in many cases in mathematics and science. By using the bundle diagrams to set limits on our choice of chunks, we see the edge of consistent reasoning at the meta-level. So, we have pushed the problem of explosion into a smaller corner than it once occupied.

Acknowledgements We are grateful to Diderik Batens, Koen Lefever, Elías Okón Gurvich, Paweł Pawłowski and Elisángela Ramírez-Cámara for the feedback offered throughout the development of this research. Special thanks are deserved to the anonymous referees for their comments and suggestions. The first author would like to acknowledge active support from the Vrije Universiteit Brussel and the Research Stays Program (PREI)-UNAM. The second author was supported by the PAPIIT Projects IA401117 "Philosophical Aspects of Contra-Classical Logics" and IA401717 "Pluralism and Normativity in Logic and Mathematics".

References

1. Abramsky, Samson. 2015. *Rui Soares Barbosa, Kohei Kishida, Raymond Lal, and Shane Mansfield*. Cohomology and Paradox: Contextuality. February.
2. Andréka, Hajnal, Németi, István, and Ahmed, Tarek Sayed. 2008. Omitting types for finite variable fragments and complete representations of algebras. *Journal of Symbolic Logic* 73 (1): 65–89. https://doi.org/10.2178/jsl/1208358743, https://projecteuclid.org/euclid.jsl/1208358743.
3. Batens, Diderik. 1992. Do we need a hierarchical model of science? *Inference, explanation, and other frustrations*, 199–215. Oakland: University of California press.
4. Benham, Richard, Chris Mortensen, and Graham Priest. 2014. Chunk and permeate III: the Dirac delta function. *Synthese* 191 (13): 3057–3062.
5. Berkeley, George. 1734. The Analyst (online edition). In: David R. Wilkins (ed.) Last accessed: 23 February 2013. http://www.maths.tcd.ie/pub/HistMath/People/Berkeley/Analyst/.
6. Brown, Bryson. 1990. How to be realistic about inconsistency in science. *Studies in History and Philosophy of Science Part A* 21 (2): 281–294.
7. Brown, Bryson. 2016. From consequence relation to reasoning strategies. *"Shut up," he explained. Essays in Honour of Peter K. Schotch*, 1–10. Milton Keynes: College Publications.
8. Brown, Bryson. 2016. On the preservation of reliability. *Logical studies of paraconsistent reasoning in science and mathematics*, 65–80., Trends in logic Cham: Springer. https://doi.org/10.1007/978-3-319-40220-8_4.
9. Brown, Bryson. 2017. Paraconsistency, pluralistic models and reasoning in climate science.
10. Brown, Bryson, and Graham Priest. 2004. Chunk and permeate, a paraconsistent inference strategy. part i: the infinitesimal calculus. *Journal of Philosophical Logic* 33 (4): 379–388. August.
11. Brown, M.Bryson, and Graham Priest. 2015. Chunk and permeate II: Bohr's hydrogen atom. *European Journal for Philosophy of Science* 5 (3): 297–314.
12. Chang, Hasok. 2014. *Is water H_2O?: evidence, realism and pluralism*, 2012th ed. Cambridge: Springer.
13. Chen, I-Tso. 2011. The liquid drop model. http://large.stanford.edu/courses/2011/ph241/chen1/.
14. Cook, Norman D. 2010. *Models of the atomic nucleus: unification through a lattice of nucleons*. Berlin: Springer.
15. Davey, Kevin. 2014. Can good science be logically inconsistent?. *Synthese* (Is Science Inconsistent? Special Issue) 191: 3009–3026.
16. Fossion, Ruben, Coster De Caroline, J.E. Garcia-Ramos, T. Werner, and Kristiaan Heyde. 2002. Nuclear binding energies: global collective structure and local shell-model correlations. *Nuclear Physics A* 697 (3–4): 703–747.

17. Fowler, A. 1913. The Spectra of Helium and Hyrogen. *Nature* 92: 95–6.
18. Friend, Michele. 2014. *Pluralism in mathematics: a new position in philosophy of mathematics*. Berlin: Springer.
19. Frisch, Mathias. 2004. Inconsistency in Classical Electrodynamics. *Philosophy of Science* 71: 525–49.
20. Frisch, Mathias. 2005. Inconsistency, Asymmetry, and Non-Locality. Oxford: Oxford University Press.
21. Friend, Michele. 2017. Inconsistency in mathematics and inconsistency in chemistry. *Humana. Mente* 32: 31–51.
22. Greiner, W., and J. A. Maruhn. 1996. Nuclear Models, Springer, Berlin.
23. Hellman, Geoffrey, and John L. Bell. 2006. Pluralism and the foundations of mathematics. *Scientific Pluralism* 19: 64–79.
24. Hempel, Carl G. 2000. *Selected philosophical essays*. New York: Cambridge University Press.
25. Imre, Lakatos. 1965. *Criticism and the Growth of Knowledge: Proceedings of the International Colloquium in the Philosophy of Science, London, 1965, Vol. 4*. Cambridge University Press, Cambridge, 1st edition edition, October 1970.
26. Krause, Decio, and Jonas RB Arenhart. 2016. *The logical foundations of scientific theories: languages, structures, and models*. Abingdon: Routledge.
27. Laudan, Larry. 1977. *Progress and its problems: towards a theory of scientific growth*, vol. 282. Berkeley: University of California Press.
28. Laudan, Larry. 1981. A problem-solving approach to scientific Progress. In: I. Hacking (ed.) Scientific Revolutions. Oxford: Oxford University Press, p. 148.
29. Jean-Marc. Lévy-Leblond. 2014. *La science est-elle universelle?* Fondation Clouste Gulbenkian. Art et Caractére.
30. Makinson, David. 2003. Bridges between classical and nonmonotonic logic. *Logic Journal of IGPL* 11 (1): 69–96.
31. Makinson, David. 2005. *Bridges from classical to nonmonotonic logic*. London: King's College Publications.
32. Margaret, Morrison. 2011. One phenomenon, many models: inconsistency and complementarity. *Studies in History and Philosophy of Science* Part A 42 (2): 342–351.
33. Martínez-Ordaz, María del Rosario. 2014. A classification for functional and inconsistent scientific theories: scope and limitations. Master Thesis. UNAM, México. In Spanish.
34. Meheus, Joke. 2002. How to reason sensibly yet naturally from inconsistencies. *Inconsistency in Science*, 151–164. Berlin: Springer.
35. Meheus, Joke. 2002. *Inconsistency in science*, vol. 2. Berlin: Springer Science & Business Media.
36. Mohamed, Elsamahi. 2005. Coherence between theories. *Canadian Journal of Philosophy* 35 (2): 331–352.
37. Morrison, Margaret. 2015. *Reconstructing reality: models, mathematics, and simulations*. Oxford: Oxford Studies in Philosophy.
38. Needham, Paul. 2015. One substance or more? *Philosophy of Chemistry*, 91–105. Berlin: Springer.
39. Peter, Vickers. 2013. *Understanding inconsistent science*. Oxford University Press.
40. Peter, Vickers. 2014. Theory flexibility and inconsistency in science. *Synthese* (Is Science Inconsistent? Special Issue) 191: 2891–2906.
41. Piazza, Mario, and Gabriele Pulcini. 2016. *Uniqueness of axiomatic extensions of cut-free classical propositional logic*. Oxford: Oxford University Press.
42. Popper, Karl. 1959. *The logic of scientific discovery*. Abingdon: Routledge.
43. Priest, Graham. 2002. Inconsistency and the empirical sciences. *Inconsistency in science*, 119–128. Berlin: Springer.
44. Priest, Graham. 2006. *Doubt truth to be a liar*. Oxford: Oxford University Press.
45. Priest, Graham. 2014. Logical pluralism: another application for chunk and permeate. *Erkenntnis* 79 (2): 331–338.

46. Priest, Graham, and Routley, Richard 1983. On Paraconsistency. Research Report 13, Logic Group, Research School of Social Sciences, Australian National University.
47. Rescher, Nicholas, and Ruth Manor. 1970. On inference from inconsistent premisses. *Theory and decision* 1 (2): 179–217.
48. Schummer, Joachim. 2015. The methodological pluralism of chemistry and its philosophical implications. *Philosophy of Chemistry*, 57–72. Berlin: Springer.
49. von Groote, H., Hilf, E. R., and K. Takahashi. 1976. New semiempirical shell correction to the droplet model. [Surface diffuseness, FORTRAN, mass formula, tables]. United States. Web. https://doi.org/10.1016/0092-640X(76)90031-0.
50. Ya Amusia, M., and Yu. Kornyushin. 2000. Application of the nuclear liquid drop model to atomic and molecular physics problems. *Contemporary Physics* 41 (4): 219–229.

What is a Paraconsistent Logic?

Eduardo Barrio, Federico Pailos and Damian Szmuc

Abstract Paraconsistent logics are logical systems that reject the classical principle, usually dubbed *Explosion*, that a contradiction implies everything. However, the received view about paraconsistency focuses only the *inferential* version of Explosion, which is concerned with *formulae*, thereby overlooking other possible accounts. In this paper, we propose to focus, additionally, on a *meta-inferential* version of Explosion, i.e. which is concerned with inferences or *sequents*. In doing so, we will offer a new characterization of paraconsistency by means of which a logic is paraconsistent if it invalidates *either* the inferential or the meta-inferential notion of Explosion. We show the non-triviality of this criterion by discussing a number of logics. On the one hand, logics which validate and invalidate both versions of Explosion, such as classical logic and Asenjo–Priest's 3-valued logic **LP**. On the other hand, logics which validate one version of Explosion but not the other, such as the substructural logics **TS** and **ST**, introduced by Malinowski and Cobreros, Egré, Ripley and van Rooij, which are obtained via Malinowski's and Frankowski's q- and p-matrices, respectively.

1 Introduction

Paraconsistent logics are logical systems that rebel against the classical principle, usually dubbed *Explosion*, that a contradiction implies everything, or that from a contradiction, everything follows.

E. Barrio (✉) · F. Pailos · D. Szmuc
IIF-SADAF, National Scientific and Technical Research Council (CONICET),
642 Bulnes, C1176ABL Buenos Aires, Argentina
e-mail: eabarrio@gmail.com

F. Pailos
e-mail: fpailos@hotmail.com

D. Szmuc
e-mail: szmucdamian@gmail.com

W. Carnielli and J. Malinowski (eds.), *Contradictions, from Consistency to Inconsistency*, Trends in Logic 47, https://doi.org/10.1007/978-3-319-98797-2_5

As Priest, Tanaka and Weber say

> The contemporary logical orthodoxy has it that, from contradictory premises, anything can be inferred (…) Inconsistency, according to received wisdom, cannot be coherently reasoned about (…) Paraconsistent logic challenges this orthodoxy. A logical consequence relation is said to be paraconsistent if it is not explosive. [24]

Similarly, in the recent book by Carnielli and Coniglio, it is said that

> Paraconsistent logics are able to deal with contradictory scenarios, avoiding triviality by means of the rejection of the Principle of Explosion. [6, p. 3]

In a nutshell, as Ripley puts it

> paraconsistency is a *non*entailment claim. [28, p. 773]

The aim of this paper is to offer a *new* characterization of what paraconsistent logics are. Our main claim will be that a logic **L** is paraconsistent if either the inferential or the meta-inferential formulation of Explosion is *invalid* in it. These two formulations of Explosion are, respectively, as follows

$$A, \neg A \Rightarrow B \qquad \frac{\Rightarrow A \quad \Rightarrow \neg A}{\Rightarrow B}$$

Where the inferential and the meta-inferential level coincide, roughly, with what are called (after Avron's work in [2]), the internal and the external consequence of a given logic.[1] Let us clarify why we take our proposal to be a non-trivial contribution to the debate about paraconsistency.

First, the received view about paraconsistency has only focused on formulations of Explosion that concern *formulae*, i.e. a formula A and its negation $\neg A$. But surely this can be taken to be a restricted point of view. In what follows we will try to broaden this conception by putting forward the aforementioned *two different formulations* of Explosion: while the former (the traditional form, that is) is concerned with formulae, the latter is concerned with *inferences* or *sequents*. Thus, the traditional conception understands Explosion as an *inference*, whereas the supplementary conception that we are trying to bring to the table *also* suggest to understand Explosion as a *meta-inference*.

Secondly, this raises the question about the possibility of finding paraconsistent logics that are so for different inferential reasons. That this possibility is real implies that our proposed criterion does not collapse with previous characterizations. In other words, it does *not* make the (in)validity of Explosion at either of these levels to *collapse* into the (in)validity at the other level. To prove this, we will offer examples of logics which validate both versions of Explosion, logics that invalidate both, and of logics that invalidate only one of them but not the other.

---

[1] In this paper we will be focusing on the inferential and the meta-inferential level, but when making our closing remarks in Sect. 5 we will point towards a plausible (although not developed here) more general conception of paraconsistency, which will require looking at *many* more levels.

To carry out our current investigation, this paper is structured as follows. In Sect. 2 we introduce the distinction between inferences and meta-inferences, along with the inferential and the meta-inferential formulations of Explosion, and our new criteria for *paraconsistency*. In Sect. 3 we present four study cases: one logic that is not paraconsistent, i.e. classical logic, and three logics that are paraconsistent: Asenjo–Priest's **LP**, and two substructural logics **TS** (a q-logic, as defined in [19], discussed by Cobreros, Ripley, Egré and van Rooij [8], Malinowski [21], and French [14]) and **ST** (a p-logic, as defined in [12], discussed by Cobreros, Ripley, Egré and van Rooij [8]). These logics are shown to be paraconsistent in different inferential ways. While **LP** invalidates both the inferential and the meta-inferential formulations of Explosion, **TS** invalidates the former but not the latter, whereas **ST** validates the former but not the latter. In Sect. 4 we provide some philosophical reflections drawn from our previous discussions, connecting our results with the debate on logical pluralism and the inferentialist stance towards the meaning of the logical connectives. Moreover, we consider three possible objections against our account and provide replies to all of them. Finally, in Sect. 5 we offer some concluding remarks, and point to some directions in which the present explorations can be further developed.

2 Different Inferential Ways of Being Paraconsistent

2.1 Inferences and Meta-Inferences

In order to understand and carry on our investigation, it will be important to have a more precise grasp of the received view about paraconsistent logics and Explosion. This view, traditionally takes paraconsistent logics as Tarskian logics and, so, we shall better understand what these are.

For the purpose of analyzing these matters, it will be useful to fix some terminology. Let $\mathscr{L}$ be a propositional language, such that $\mathbf{FOR}(\mathscr{L})$ is the absolutely free algebra of formulae of $\mathscr{L}$, whose universe we denote by $FOR(\mathscr{L})$.

Definition 1 A *Tarskian* consequence relation over a propositional language $\mathscr{L}$ is a relation $\vDash\ \subseteq \wp(FOR(\mathscr{L})) \times FOR(\mathscr{L})$ obeying the following conditions for all $A \in FOR(\mathscr{L})$ and for all $\Gamma, \Delta \subseteq FOR(\mathscr{L})$:

1. $\Gamma \vDash A$ if $A \in \Gamma$ (Reflexivity)
2. If $\Gamma \vDash A$ and $\Gamma \subseteq \Gamma'$, then $\Gamma' \vDash A$ (Monotonicity)
3. If $\Delta \vDash A$ and $\Gamma \vDash B$ for every $B \in \Delta$, then $\Gamma \vDash A$ (Cut)

Additionally, a (Tarskian) consequence relation $\vDash$ is substitution-invariant whenever if $\Gamma \vDash A$, and σ is a substitution on $FOR(\mathscr{L})$, then $\{\sigma(B) \mid B \in \Gamma\} \vDash \sigma(A)$.

Definition 2 A Tarskian logic over a propositional language $\mathscr{L}$ is an ordered pair $(FOR(\mathscr{L}), \vDash)$, where $\vDash$ is a substitution-invariant Tarskian consequence relation.

Throughout the years many scholars have argued that the Tarskian conception of logic is quite narrow. For example, Shoesmith and Smiley [30], Avron [3] and Scott [29] claimed that the Tarskian account should be generalized to a logic having multiple consequences; and Avron [3] and Gabbay [15] have argued that the condition of Monotonicity should be relaxed; whereas it can be inferred that, derivatively, Malinowski [19] and Frankowski [12] argued for a generalization or liberalization which allows logics to drop Reflexivity and/or Cut.

These modifications, in turn, can be made sense of by noticing a shift in the nature of the collection of *formulae* featured in the consequence relation. Thus, for example, instead of treating logical consequence to hold between (sets of) formulae, it may hold between labelled formulae, sequences of formulae (where order matters), multisets of formulae (where repetition matters), etc. Interestingly, many of these approaches invalidate Explosion, regarded as an inference that relates collections (sets, sequences, multisets, etc.) of *formulae*. But none of the aforementioned alternatives proposed explicitly to move from logical consequence as a relation conceived between collections of *formulae* to a relation conceived between collections of some *other entities*. Therefore, none of these alternatives proposed explicitly to change from focusing on Explosion as an inference that relates formulae to an inference that relates other entities.

However, some other approaches did. That is the case of Avron in [2], first, and Blok and Jónsson in [5], second, which discuss a generalization of the Tarksian account that allows to move to logical consequence relations that do not hold only between collections of formulae, but between objects of other nature.

Definition 3 An inference or *sequent* on $\mathscr{L}$ is an ordered pair (Γ, A), where $\Gamma \subseteq FOR(\mathscr{L})$ and $A \in FOR(\mathscr{L})$ (written $\Gamma \Rightarrow A$). $SEQ^0(\mathscr{L})$ is the set of all inferences or sequents on $\mathscr{L}$.

Definition 4 ([11]) A meta-inference or *meta-sequent* on $\mathscr{L}$ is an ordered pair (Γ, A), where $\Gamma \subseteq SEQ^0(\mathscr{L})$ and $A \in SEQ^0(\mathscr{L})$ (written $\Gamma \Rightarrow^1 A$). $SEQ^1(\mathscr{L})$ is the set of all meta-inferences or meta-sequents on $\mathscr{L}$.

We will say, accordingly, that from the following the one on the left is an inference, whereas the one on the right is a meta-inference

$$A, B \Rightarrow A \wedge B \qquad \qquad \frac{\Rightarrow A \quad \Rightarrow B}{\Rightarrow A \wedge B}$$

and, indeed, according to the following definitions adapted from Avron [2], both are valid in e.g. Gentzen's sequent calculus **LK** for classical logic—as we shall see next, when we define the corresponding notions of validity.

Now, going back to the proposed shifts from the ontology of the Tarskian account of logical consequence, Avron suggested in [2] that the idea that logical consequence can be said to hold of *relata* other than formulae is very reasonable to those used to *sequent calculus*—and, most prominently, with substructural sequent calculi.

For Avron there are two different notions of logical consequence for a given sequent calculus **S**: the *internal* and the *external* notion of logical consequence. In our work, however, instead of referring to these relations as internal and external, we will refer to these levels, respectively, as the *inferential* and the *meta-inferential*, characterized such that

- A follows inferentially from Γ in **S** (written $\vdash_{\mathbf{S}} \Gamma \Rightarrow A$) whenever $\Gamma \Rightarrow A$ is a provable sequent of the calculus **S**. In such a case we will say that the inference from Γ to A is **S**-*valid*.

This relation is concerned with which *formulae* follow from which (collection of) formulae, given the rules of the calculus—i.e. which *sequents* follow, given the axioms and rules of the calculus.

- A follows meta-inferentially from Γ in **S** (written $\vdash_{\mathbf{S}} \Gamma \Rightarrow^1 A$) whenever $\Rightarrow A$ is provable in the calculus that results from the addition to **S** of all the sequents $\Rightarrow B$ (for B in Γ) as initial sequents or axioms.[2] In such a case we will say that the meta-inference from Γ to A is **S**-*valid*.

This means that this relation is concerned with which sequents follow from which (set of) sequents, given the axioms and rules of the calculus.

That these relations are different can be easily exemplified by the fact, nicely noticed by Mares and Paoli in [22], that if a sequent calculus **S** has no Weakening rules, then

$$\nvdash_{\mathbf{S}} A, B \Rightarrow A \qquad \text{although} \qquad \vdash_{\mathbf{S}} A, B \Rightarrow^1 A$$

Finally, notice also in passing that re-writing the meta-inference

$$\frac{\Rightarrow A \quad \Rightarrow \neg A}{\Rightarrow B}$$

with the aid of the previous notation, gives us as a result

$$\{\Rightarrow A, \Rightarrow \neg A\} \Rightarrow^1 \Rightarrow B$$

which, for matters of readability, we will write as

$$A, \neg A \Rightarrow^1 B$$

reinforcing, thereby, the idea that we are dealing with nothing more than yet another formulation of Explosion. Something that we will argue for explicitly in the next section.

[2] We shall notice that in [2] Avron takes this definition to, additionally, require that Cut is taken as a primitive rule. Something that—for the sake of generality—we do not demand here.

2.2 Explosion, Revisited

Explosion, so to speak, comes in different flavors. Many rules, meta-rules and principles are dubbed with that name. Nevertheless, there are what seems to be some central or *essential* features that every one of them share, and indeed it is that they embody the idea that *contradiction* equals *triviality*. This is traditionally understood, in terms of the received view about paraconsistency as saying that

An INFERENCE *with an inconsistent premise set implies any conclusion*

As is well-known, inconsistent premise sets for inferences are sets that include (some instance of) the (schematic) *formulae* A and $\neg A$. So, along these lines, Explosion is without any surprise taken to be the inference

$$A, \neg A \Rightarrow B$$

The question is now, how to adapt this idea to the case of meta-inferences. For us, the most reasonable take is to say that

A META- INFERENCE *with an inconsistent premise set implies any conclusion*

But, now, for meta-inferences, we must keep in mind that premise sets and conclusions are formed with sequents. Thus, we must define what an inconsistent premise set for meta-inferences is. We take these to be sets that include (some instance of) the (schematic) *sequents* $\Rightarrow A$ and $\Rightarrow \neg A$. That this is, in fact, a right way to understand an inconsistent sequent set can be argued for by looking at e.g. the definition of an inconsistent belief set (cf. [17]). Along these lines, Explosion is without any surprise taken to be the meta-inference

$$\frac{\Rightarrow A \quad \Rightarrow \neg A}{\Rightarrow B}$$

Furthermore, we are in good company in claiming that these are in fact two versions or formulations of Explosion, one as an inference and the other as a meta-inference. For Lloyd Humberstone, in his reference book *The Connectives* says in [18, p. 118–119] that $A, \neg A \Rightarrow B$ is a sequent, i.e. an *inferential* form of Explosion, whereas the following is a rule, i.e. a *meta-inferential* form of Explosion.

$$\frac{\Gamma \Rightarrow A \quad \Delta \Rightarrow \neg A}{\Gamma, \Delta \Rightarrow B}$$

By letting Γ and Δ be empty, Humberstone proposed form collapses with ours.[3]

To conclude, let us rephrase in a more formal manner the main claim of this paper

[3]Let us notice, additionally, that Humberstone calls Ex Falso Quodlibet what we call Explosion, but this is just a terminological and non-substantial issue.

$$A \text{ LOGIC } \mathbf{L} \textit{ is } \text{PARACONSISTENT } \textit{if} \begin{cases} \textit{either} & A, \neg A \Rightarrow B \textit{ is invalid in } \mathbf{L} \\ \textit{or} & A, \neg A \Rightarrow^1 B \textit{ is invalid in } \mathbf{L} \end{cases}$$

Additionally, let us highlight that we are not claiming these two are the *only* dresses that Explosion can use. If a logic invalidates either of the previous formulations of Explosion, we will say that it is paraconsistent, although it *need not* invalidate either to be so, for it may invalidate some other formulation(s) of Explosion. For example, Explosion is sometimes formulated with the help of conjunction. Exploring 'conjunctive' versions of Explosion (both at the inferential and the meta-inferential level) is no doubt an interesting task, one which for matters of space we decided not to tackle here. In other words, we are not proposing a necessary, but a new *sufficient* condition for logics to be paraconsistent.

An additional caveat, which echoes the well-known reservations expressed by Igor Urbas in [34], should be mentioned concerning this characterization. In his work, Urbas points out that the traditional definition of paraconsistency in terms of invalidating the inferential form of Explosion counts as paraconsistent some logics "which satisfy the letter of [this criterion] while brazenly flouting its spirit" [34, p. 345]. For instance, Johansson's Minimal Logic invalidates Explosion, while validating the scheme $A, \neg A \Rightarrow \neg B$, for arbitrary formulae B. Furthermore, as highlighted by an anonymous referee, this logic will also invalidate the meta-inferential formulation of Explosion, while still validating the scheme $A, \neg A \Rightarrow^1 \neg B$, for arbitrary formulae B. Thus, these considerations lead us to note that these pathological cases can—and probably should—be exempted from the definition.

In what follows we will compare different cases of different logics, showing that all of them take a distinctive stance with regard to the valid or invalid character of the above portrayed inferential and meta-inferential versions of Explosion.

3 Study Cases

To accomplish our task in this section, we will divide these systems in two groups. The first group will be composed of *matrix logics* and will include a logic that is *not* paraconsistent at any level, i.e. classical logic **CL**, and a logic that is paraconsistent *both* at the inferential and the meta-inferential level, i.e. Asenjo–Priest's 3-valued logic **LP** from [1, 23]. The second group will be composed of a q-matrix logic,[4] i.e. the logic **TS**, and a p-matrix logic,[5] i.e. the logic **ST**, both due to Cobreros, Ripley, Egré and van Rooij in [8], the former being also discussed by Malinowski in [21]. These logics will be shown to be, respectively, paraconsistent at the inferential but not the meta-inferential level, and paraconsistent at the meta-inferential but not the inferential level.

[4] q-consequence relations and q-matrices were introduced by Grzegorz Malinowski in [19].

[5] p-consequence relations and p-matrices were introduced by Szymon Frankowski in [12].

3.1 Matrix Logics

Definition 5 For $\mathscr{L}$ a propositional language, an $\mathscr{L}$-matrix is a structure $\mathscr{M} = \langle \mathscr{V}, \mathscr{D}, \mathscr{O} \rangle$, such that $\langle \mathscr{V}, \mathscr{O} \rangle$ is an algebra of the same similarity type as $\mathscr{L}$, with universe $\mathscr{V}$ and a set of operations $\mathscr{O}$, and $\mathscr{D} \subseteq \mathscr{V}$.

Notice, in the first place, that the set $\mathscr{O}$ includes for every n-ary connective $\diamond$ in the language $\mathscr{L}$, a corresponding n-ary truth-function $f^{\diamond}_{\mathscr{M}} : \mathscr{V}^n \longrightarrow \mathscr{V}$. With regard to these, when context allows it, we will sometimes identify the connectives themselves (which are linguistic items), with their corresponding truth-functions in a given matrix. In the second place, notice that typically, when dealing with non-classical logics, the set $\mathscr{V}$ is taken to be a superset of $\{\mathbf{t}, \mathbf{f}\}$.

Definition 6 For $\mathscr{M}$ an $\mathscr{L}$-matrix (respectively, an $\mathscr{L}$-q-matrix or an $\mathscr{L}$-$p$matrix), an $\mathscr{M}$-valuation v is an homomorphism from $FOR(\mathscr{L})$ to $\mathscr{V}$, for which we denote by $v[\Gamma]$ the set $\{v(B) \mid B \in \Gamma\}$, i.e. the image of v under Γ.

Of interest are two-valued classical logic **CL** (which we do not bother to present here due to the fact that it is perhaps the best known matrix logic), and the Asenjo–Priest's 3-valued logic **LP**, which is defined based on the 3-element Kleene algebra.

Definition 7 The 3-element Kleene algebra is the structure

$$\mathbf{K} = \langle \{\mathbf{t}, \mathbf{i}, \mathbf{f}\}, \{f^{\neg}_{\mathbf{K}}, f^{\wedge}_{\mathbf{K}}, f^{\vee}_{\mathbf{K}}\} \rangle$$

where the functions $f^{\neg}_{\mathbf{K}}, f^{\wedge}_{\mathbf{K}}, f^{\vee}_{\mathbf{K}}$ are as follows

	$f^{\neg}_{\mathbf{K}}$
t	**f**
i	**i**
f	**t**

$f^{\wedge}_{\mathbf{K}}$	**t**	**i**	**f**
t	**t**	**i**	**f**
i	**i**	**i**	**f**
f	**f**	**f**	**f**

$f^{\vee}_{\mathbf{K}}$	**t**	**i**	**f**
t	**t**	**t**	**t**
i	**t**	**i**	**i**
f	**t**	**i**	**f**

Definition 8 ([1, 23]) A 3-valued **LP**-matrix is a structure

$$\mathscr{M}_{\mathbf{LP}} = \langle \{\mathbf{t}, \mathbf{i}, \mathbf{f}\}, \{\mathbf{t}, \mathbf{i}\}, \{f^{\neg}_{\mathbf{K}}, f^{\wedge}_{\mathbf{K}}, f^{\vee}_{\mathbf{K}}\} \rangle$$

such that $\langle \{\mathbf{t}, \mathbf{i}, \mathbf{f}\}, \{f^{\neg}_{\mathbf{K}}, f^{\wedge}_{\mathbf{K}}, f^{\vee}_{\mathbf{K}}\} \rangle$ is the 3-element Kleene algebra.

With these definitions we are now in a position to ask which formulations of Explosion are valid, and which are invalid in classical logic and in **LP**. However, for this question to be meaningful, it is necessary to clarify how matrix logics validate or invalidate both inferences and meta-inferences. Notice that, below, $\vDash_{\mathscr{M}}$ is a substitution-invariant Tarskian consequence relation over $\mathscr{L}$, whence $(FOR(\mathscr{L}), \vDash_{\mathscr{M}})$ is a Tarskian logic. In addition to that, when some logic **L** is induced by a matrix $\mathscr{M}$, we may interchangeably refer to $\vDash_{\mathscr{M}}$ as $\vDash_{\mathbf{L}}$.

Definition 9 For $\mathcal{M}$ a matrix, an $\mathcal{M}$-valuation v satisfies a sequent or inference $\Gamma \Rightarrow A$ (written $v \vDash_{\mathcal{M}} \Gamma \Rightarrow A$) iff if $v[\Gamma] \subseteq \mathcal{D}$, then $v(A) \in \mathcal{D}$. A sequent or inference $\Gamma \Rightarrow A$ is $\mathcal{M}$-valid (written $\vDash_{\mathcal{M}} \Gamma \Rightarrow A$) iff $v \vDash_{\mathcal{M}} \Gamma \Rightarrow A$, for all $\mathcal{M}$-valuations v.

Definition 10 For $\mathcal{M}$ a matrix, an $\mathcal{M}$-valuation v satisfies a meta-sequent or meta-inference $\Gamma \Rightarrow^1 A$ (written $v \vDash_{\mathcal{M}} \Gamma \Rightarrow^1 A$) iff if $v \vDash_{\mathcal{M}} B$, for all $B \in \Gamma$, then $v \vDash_{\mathcal{M}} A$. A meta-sequent or meta-inference $\Gamma \Rightarrow^1 A$ is $\mathcal{M}$-valid (written $\vDash_{\mathcal{M}} \Gamma \Rightarrow^1 A$) iff if $v \vDash_{\mathcal{M}} B$, for all $B \in \Gamma$, then $v \vDash_{\mathcal{M}} A$, for all $\mathcal{M}$-valuations v.

Recall that in the last definition e.g. B stands for a sequent, i.e. an object of the form $\Sigma \Rightarrow C$, and therefore e.g. $v \vDash_{\mathcal{M}} B$ should be read as $v \vDash_{\mathcal{M}} \Sigma \Rightarrow C$. Accordingly, when Σ is empty, it should be read as $\vDash_{\mathcal{M}} \emptyset \Rightarrow C$, which for matters of readability we write as $\vDash_{\mathcal{M}} \Rightarrow C$.

Given these definitions it is easy to observe the following facts.

Fact 3.1 *Classical logic* **CL** *validates both the inferential and the meta-inferential formulation of Explosion, i.e.* $\vDash_{\mathbf{CL}} A, \neg A \Rightarrow B$ *and* $\vDash_{\mathbf{CL}} A, \neg A \Rightarrow^1 B$.

Proof These two facts are straightforwardly verified by noticing that there is no **CL**-valuation v such that $v(\{A, \neg A\}) \subseteq \{\mathbf{t}\}$, i.e. that there is no **CL**-valuation v such that $v \vDash_{\mathbf{CL}} \Rightarrow A$ and $v \vDash_{\mathbf{CL}} \Rightarrow \neg A$. From this we infer, on the one hand, that there is no **CL**-valuation v such that $v(\{A, \neg A\}) \subseteq \{\mathbf{t}\}$ and $v(B) \notin \{\mathbf{t}\}$, whence $\vDash_{\mathbf{CL}} A, \neg A \Rightarrow B$. And, on the other hand, that there is no **CL**-valuation v such that $v \vDash_{\mathbf{CL}} \Rightarrow A$ and $v \vDash_{\mathbf{CL}} \Rightarrow \neg A$ and $v \nvDash_{\mathbf{CL}} \Rightarrow B$, whence $\vDash_{\mathbf{CL}} A, \neg A \Rightarrow^1 B$. □

Fact 3.2 *The logic* **LP** *invalidates both the inferential and the meta-inferential formulation of Explosion, i.e.* $\nvDash_{\mathbf{LP}} A, \neg A \Rightarrow B$ *and* $\nvDash_{\mathbf{LP}} A, \neg A \Rightarrow^1 B$.

Proof To prove this facts, it is routine to construct an **LP**-valuation v such that $v(A) = v(\neg A) = \mathbf{i}$, while $v(B) = \mathbf{f}$. From this we infer, on the one hand, that v is a valuation such that $v(\{A, \neg A\}) \subseteq \{\mathbf{t}, \mathbf{i}\}$ and $v(B) \notin \{\mathbf{t}, \mathbf{i}\}$, whence $\nvDash_{\mathbf{LP}} A, \neg A \Rightarrow B$. On the other hand, we infer that v is a valuation such that $v \vDash_{\mathbf{LP}} \Rightarrow A$ and $v \vDash_{\mathbf{LP}} \Rightarrow \neg A$, while $v \nvDash_{\mathbf{LP}} \Rightarrow B$, whence we conclude that $\nvDash_{\mathbf{LP}} A, \neg A \Rightarrow^1 B$. □

3.2 q-Matrix Logics and p-Matrix Logics

Two interesting generalizations of Tarskian consequence relations appeared in the last two decades, the notion of q-consequence relation, due to Malinowski [19] and the notion of p-consequence relation, due to Frankowski [12]. As Wansing and Shramko clearly explain in [31], the corresponding relation of q-logic is devised to qualify as valid derivations of true sentences from non-refuted premises (understood as hypotheses), whereas the notion of p-logic is devised to qualify as valid derivations of conclusions whose degree of strength (understood as the conviction in its truth) is smaller than that of the premises. We define these notions formally as follows.

Definition 11 ([19]) A q-consequence relation over a propositional language $\mathscr{L}$ is a relation $\vDash\, \subseteq \wp(FOR(\mathscr{L})) \times FOR(\mathscr{L})$ obeying the following conditions for all $A \in FOR(\mathscr{L})$ and for all $\Gamma, \Delta \subseteq FOR(\mathscr{L})$:

1. If $\Gamma \vDash A$ and $\Gamma \subseteq \Gamma'$, then $\Gamma' \vDash A$ (Monotonicity)
2. $\Gamma \cup \{B \mid \Gamma \vDash B\} \vDash A$ iff $\Gamma \vDash A$ (Quasi-closure)

Definition 12 ([19]) A q-logic over a propositional language $\mathscr{L}$ is an ordered pair $(FOR(\mathscr{L}), \vDash)$, where $\vDash$ is a substitution-invariant q-consequence relation.

Definition 13 ([12]) A p-consequence relation over a propositional language $\mathscr{L}$ is a relation $\vDash\, \subseteq \wp(FOR(\mathscr{L})) \times FOR(\mathscr{L})$ obeying the following conditions for all $A \in FOR(\mathscr{L})$ and for all $\Gamma, \Delta \subseteq FOR(\mathscr{L})$:

1. $\Gamma \vDash A$ if $A \in \Gamma$ (Reflexivity)
2. If $\Gamma \vDash A$ and $\Gamma \subseteq \Gamma'$, then $\Gamma' \vDash A$ (Monotonicity)

Definition 14 ([12]) A p-logic over a propositional language $\mathscr{L}$ is an ordered pair $(FOR(\mathscr{L}), \vDash)$, where $\vDash$ is a substitution-invariant p-consequence relation.

Notice, moreover, that q-logics fail to validate Reflexivity, while p-logics fail to validate Cut and, thus, are both non-Tarskian or *substructural* logics.

Semantically speaking, q-logics and p-logics can be obtained from structures called, respectively, q-matrices and p-matrices, by similar means than Tarskian logics are obtained from regular matrices. Whence, we may refer to them as q-matrix logics and p-matrix logics.

Definition 15 ([19]) For $\mathscr{L}$ a propositional language, an $\mathscr{L}$-q-matrix is a structure $\langle \mathscr{V}, \mathscr{D}^+, \mathscr{D}^-, \mathscr{O} \rangle$, such that $\langle \mathscr{V}, \mathscr{O} \rangle$ is an algebra of the same similarity type as $\mathscr{L}$, with universe $\mathscr{V}$ and a set of operations $\mathscr{O}$, where $\mathscr{D}^+, \mathscr{D}^- \subseteq \mathscr{V}$ and $\mathscr{D}^+ \cap \mathscr{D}^- = \emptyset$.

Definition 16 ([13]) For $\mathscr{L}$ a propositional language, an $\mathscr{L}$-p-matrix is a structure $\langle \mathscr{V}, \mathscr{D}^+, \mathscr{D}^-, \mathscr{O} \rangle$, such that $\langle \mathscr{V}, \mathscr{O} \rangle$ is an algebra of the same similarity type as $\mathscr{L}$, with universe $\mathscr{V}$ and a set of operations $\mathscr{O}$, where $\mathscr{D}^+, \mathscr{D}^- \subseteq \mathscr{V}$ and $\mathscr{D}^+ \subseteq \mathscr{D}^-$.

A word on how q- and p-matrices generalize the usual notion of a logical matrix is in order. In a usual logical matrix $\langle \mathscr{V}, \mathscr{D}, \mathscr{O} \rangle$ the truth-values of the matrix, i.e. the elements of $\mathscr{V}$, are presented in a *dichotomized* way. By this we mean that they either belong to $\mathscr{D}$—and, hence, are designated—or they belong to $\mathscr{V} \setminus \mathscr{D}$—and, hence, are anti-designated.

Contrary to this, q- and p-matrices start from a non-dichotomized classification of the truth-values of the given matrix—i.e. the members of $\mathscr{V}$—letting them belong to two sets, which we here call $\mathscr{D}^+$ and $\mathscr{D}^-$.[6] We will, then, allow these sets to be *jointly*

[6] Here we will adopt this terminology—i.e. talk of $\mathscr{D}^+$ and $\mathscr{D}^-$—introduced by [32], emphasizing that we will take q- and p-logics to be induced by different type of structures, i.e. respectively q- and p-matrices. In this vein, what will be distinctive of these type of structures will be the properties of the sets $\mathscr{D}^+$ and $\mathscr{D}^-$, as detailed in Definitions 15 and 16, respectively.

non-exhaustive and *mutually non-exclusive*. Paradigmatically, the first note of this generalization is associated with q-matrices, where it is allowed that $\mathscr{D}^+ \cup \mathscr{D}^- \neq \mathscr{V}$ (see e.g. [20, p. 12]). Analogously, the second note of this generalization is associated with p-matrices, where it is allowed that $\mathscr{D}^+ \cap \mathscr{D}^- \neq \emptyset$ (see e.g. [12, p. 45]).

From a purely abstract point of view, the sets $\mathscr{D}^+$ and $\mathscr{D}^-$ need not be attached any particular philosophical interpretation and, thus, the symbols $+$ and $-$ are taken by us to be *arbitrary*. Notwithstanding this, e.g. in the context of Malinowski's discussion of q-matrices, they are usually taken to represent, respectively, the set of accepted and rejected elements (see [21]). Whereas, in the context of Frankowski's discussion of p-matrices they are usually taken to represent, respectively, the set of values representing the degree of strength of the premises and the set of values representing the degree of strength of the conclusion (see [12]). Furthermore, in the context of Wansing and Shramko's discussion of q-matrices, those truth-values belonging to $\mathscr{D}^+$ are identified as representatives of a generalized notion of *truth* and those truth-values belonging to $\mathscr{D}^-$ as representatives of a generalized notion of *falsity* (see [32, p. 195]).

There are two 3-valued q- and p-matrix logics associated to the 3-element Kleene algebra that are discussed in the literature, which we would like to present in connection to our ongoing investigation: the logic **TS** and the logic **ST**.

Definition 17 ([8, 21]) A 3-valued **TS**-matrix is a q-matrix

$$\mathscr{M}_{\mathbf{TS}} = \langle \{\mathbf{t}, \mathbf{i}, \mathbf{f}\}, \{\mathbf{t}\}, \{\mathbf{f}\}, \{f_{\mathbf{K}}^{\neg}, f_{\mathbf{K}}^{\wedge}, f_{\mathbf{K}}^{\vee}\}\rangle$$

such that $\langle \{\mathbf{t}, \mathbf{i}, \mathbf{f}\}, \{f_{\mathbf{K}}^{\neg}, f_{\mathbf{K}}^{\wedge}, f_{\mathbf{K}}^{\vee}\}\rangle$ is the 3-element Kleene algebra.

Definition 18 ([8]) A 3-valued **ST**-matrix is a p-matrix

$$\mathscr{M}_{\mathbf{ST}} = \langle \{\mathbf{t}, \mathbf{i}, \mathbf{f}\}, \{\mathbf{t}\}, \{\mathbf{t}, \mathbf{i}\}, \{f_{\mathbf{K}}^{\neg}, f_{\mathbf{K}}^{\wedge}, f_{\mathbf{K}}^{\vee}\}\rangle$$

such that $\langle \{\mathbf{t}, \mathbf{i}, \mathbf{f}\}, \{f_{\mathbf{K}}^{\neg}, f_{\mathbf{K}}^{\wedge}, f_{\mathbf{K}}^{\vee}\}\rangle$ is the 3-element Kleene algebra.

The former is discussed by e.g. Cobreros, Ripley, Egré and van Rooij in [8], and also by Chemla, Egré and Spector in [7] in the context of the more general discussion of what represents a 'respectable' consequence relation between formulae. Moreover, it was also discussed by Grzegorz Malinowski in [21] as a tool to model empirical inference with the aid of the 3-valued Kleene algebra, and more recently was also stressed by Rohan French in [14], in connection with the paradoxes of self-reference.

The latter is discussed by Cobreros, Ripley, Egré and van Rooij in several papers (among them [8, 9, 26, 27]), with the aim of solving the riddles raised by paradoxical phenomena, vagueness, and much more. It must be pointed out that it was also entertained by Girard in [16] as a 3-valued interpretation of the sequent calculus **LK** for classical propositional logic, without the Cut rule.

Once more, equipped with these definitions we are now in a position to ask which formulations of Explosion are valid, and which are invalid in the q-matrix logic **TS** and the p-matrix logic **ST**. Yet again, for this question to be meaningful, it is necessary to clarify how q- and p-matrix logics validate or invalidate both inferences and meta-inferences—following e.g. [12] and [32, p. 196]. Notice that, below, $\vDash_{\mathscr{M}}$ is a substitution-invariant q-consequence (respectively, p-consequence) relation, whence $(FOR(\mathscr{L}), \vDash_{\mathscr{M}})$ is a q-logic (respectively, a p-logic). In addition to that, when some q- or p-logic **L** is induced by, respectively, a q- or p- matrix $\mathscr{M}$, we may interchangeably refer to $\vDash_{\mathscr{M}}$ as $\vDash_{\mathbf{L}}$.

Definition 19 For $\mathscr{M}$ a q-matrix, an $\mathscr{M}$-valuation v satisfies a sequent or inference $\Gamma \Rightarrow A$ (written $v \vDash_{\mathscr{M}} \Gamma \Rightarrow A$) iff if $v[\Gamma] \cap \mathscr{D}^- = \emptyset$, then $v(A) \in \mathscr{D}^+$. For $\mathscr{M}$ a p-matrix, an $\mathscr{M}$-valuation v satisfies a sequent or inference $\Gamma \Rightarrow A$ (written $v \vDash_{\mathscr{M}} \Gamma \Rightarrow A$) iff if $v[\Gamma] \subseteq \mathscr{D}^+$, then $v(A) \in \mathscr{D}^-$.[7] For $\mathscr{M}$ a q-matrix or p-matrix, a sequent or inference $\Gamma \Rightarrow A$ is $\mathscr{M}$-valid (written $\vDash_{\mathscr{M}} \Gamma \Rightarrow A$) iff $v \vDash_{\mathscr{M}} \Gamma \Rightarrow A$, for all $\mathscr{M}$-valuations v.

Definition 20 For $\mathscr{M}$ a q-matrix or p-matrix, an $\mathscr{M}$-valuation v satisfies a meta-sequent or meta-inference $\Gamma \Rightarrow^1 A$ (written $v \vDash_{\mathscr{M}} \Gamma \Rightarrow^1 A$) iff if $v \vDash_{\mathscr{M}} B$, for all $B \in \Gamma$, then $v \vDash_{\mathscr{M}} A$. A meta-sequent or meta-inference $\Gamma \Rightarrow^1 A$ is $\mathscr{M}$-valid (written $\vDash_{\mathscr{M}} \Gamma \Rightarrow^1 A$) iff if $v \vDash_{\mathscr{M}} B$, for all $B \in \Gamma$, then $v \vDash_{\mathscr{M}} A$, for all $\mathscr{M}$-valuations v.

From these definitions the following facts follow.

Fact 3.3 ([8]) **TS** *is a non-reflexive, and thus a substructural, logic.*

Fact 3.4 ([8]) **ST** *is a non-transitive, and thus a substructural, logic.*

Fact 3.5 *The logic* **TS** *invalidates the inferential formulation of Explosion, i.e.* $\nvDash_{\mathbf{TS}} A, \neg A \Rightarrow B$*, but it validates the meta-inferential formulation of Explosion, i.e.* $\vDash_{\mathbf{TS}} A, \neg A \Rightarrow^1 B$*.*

Proof To prove that $\nvDash_{\mathbf{TS}} A, \neg A \Rightarrow B$ construct a **TS**-valuation v such that $v(A) = v(\neg A) = \mathbf{i}$, i.e. $v(\{A, \neg A\}) = \{\mathbf{i}\}$, while $v(B) = \mathbf{f}$. From this we infer that v is a valuation such that $v(\{A, \neg A\}) \cap \{\mathbf{f}\} = \emptyset$ and $v(B) \notin \{\mathbf{t}\}$, whence $\nvDash_{\mathbf{TS}} A, \neg A \Rightarrow B$.

To prove that $\vDash_{\mathbf{TS}} A, \neg A \Rightarrow^1 B$, suppose for *reductio* that $\nvDash_{\mathbf{TS}} A, \neg A \Rightarrow^1 B$. Then, there should be a **TS**-valuation v, such that $v \vDash_{\mathbf{TS}} \Rightarrow A$ and $v \vDash_{\mathbf{TS}} \Rightarrow \neg A$, while $v \nvDash_{\mathbf{TS}} \Rightarrow B$. Such a valuation will require that $v(A) = \mathbf{t} = v(\neg A)$, which is impossible. Whence, we conclude $\vDash_{\mathbf{TS}} A, \neg A \Rightarrow^1 B$. □

[7] Notice that this definition takes a p-logic to be induced by a p-matrix $\langle \mathscr{V}, \mathscr{D}^+, \mathscr{D}^-, \mathscr{O} \rangle$ where it is assumed that $\mathscr{D}^+ \subseteq \mathscr{D}^-$, whence this last clause reads: "if $v[\Gamma] \subseteq \mathscr{D}^+$, then $v(A) \in \mathscr{D}^-$". Now, as remarked by an anonymous referee, if the same p-logic is taken to be induced by a q-matrix $\langle \mathscr{V}, \mathscr{D}^+, \mathscr{D}^-, \mathscr{O} \rangle$ where it is assumed that $\mathscr{D}^+ \cap \mathscr{D}^- = \emptyset$—as is done e.g. in [32, p. 210]—then this last clause should read: "if $v[\Gamma] \subseteq \mathscr{D}^+$, then $v(A) \notin \mathscr{D}^-$".

These considerations highlight that if the sets $\mathscr{D}^+$ and $\mathscr{D}^-$ of a q-matrix are taken to, respectively, represent a generalized notion of *truth* and a generalized notion of *falsity*—as in [32]—then with regard to valuations on the 3-element Kleene algebra, **TS** and **ST** can be interpreted as follows. **TS** consequence can be understood as requiring that for all valuations, if the premises are non-false, then the conclusion is true; whereas **ST** consequence can be understood as requiring that for all valuations, if the premises are true, then the conclusion is non-false.

Fact 3.6 ([4]) *The logic* **ST** *validates the inferential formulation of Explosion, i.e.* $\vDash_{\mathbf{ST}} A, \neg A \Rightarrow B$*, but it invalidates the meta-inferential formulation of Explosion, i.e.* $\nvDash_{\mathbf{ST}} A, \neg A \Rightarrow^1 B$.

Proof To prove that $\vDash_{\mathbf{ST}} A, \neg A \Rightarrow B$, suppose for *reductio* that $\nvDash_{\mathbf{ST}} A, \neg A \Rightarrow B$. Then, there should be an **ST**-valuation v, such that $v(\{A, \neg A\}) \subseteq \{\mathbf{t}\}$, while $v(B) \notin \{\mathbf{t}, \mathbf{i}\}$. Such a valuation will require that $v(A) = \mathbf{t} = v(\neg A)$, which is impossible. Whence, we conclude $\vDash_{\mathbf{ST}} A, \neg A \Rightarrow B$.

To prove that $\nvDash_{\mathbf{ST}} A, \neg A \Rightarrow^1 B$ construct an **ST**-valuation v such that $v(A) = v(\neg A) = \mathbf{i}$, i.e. $v(\{A, \neg A\}) = \{\mathbf{i}\}$, while $v(B) = \mathbf{f}$. From this we infer that v is a valuation such that $v \vDash_{\mathbf{ST}} \Rightarrow A$ and $v \vDash_{\mathbf{ST}} \Rightarrow \neg A$, while $v \nvDash_{\mathbf{ST}} \Rightarrow B$, whence $\nvDash_{\mathbf{ST}} A, \neg A \Rightarrow^1 B$. □

Before moving on, it might be worth noticing—as pointed out by an anonymous referee—that according to **TS** and **ST** the meta-inferential formulation of Explosion is closely related to a restricted form of Cut. To be more precise, given these systems validate e.g. the rule of right Weakening [*WR*] and also the left introduction rule for negation [$\neg L$], it is true that the meta-inferential formulation of Explosion is *equivalent* to a restricted form of Cut—indeed, of both the additive [Cut^{A}] or the multiplicative [Cut^{M}] version of Cut[8]—where the side formulae are *empty*.[9]

We can, in fact, provide more general facts from which the previous can be seen as corollaries. We do think that, nevertheless, giving the proper counterexamples for the particular cases above is illustrative, as these logics are not so commonly mentioned in the literature about paraconsistent logics.

Fact 3.7 **TS** *has no valid inferences or sequents.*

Proof Consider an arbitrary inference or sequent $\Gamma \Rightarrow A$, and consider a **TS**-valuation v, such that v assigns the value $\mathbf{i}$ to every propositional variable of $\mathscr{L}$. Given **TS** is a q-matrix based on the 3-valued Kleene algebra, it is easy to see by looking at the operations of the algebra that if every propositional variable p is such that $v(p) = \mathbf{i}$, then every formula C is such that $v(C) = \mathbf{i}$, and in particular for every $B \in \Gamma$, $v(B) = \mathbf{i}$. Now, it only remains to notice that v is a **TS**-valuation such that $v[\Gamma] \cap \{\mathbf{f}\} = \emptyset$, but $v(A) \notin \{\mathbf{t}\}$, whence $\nvDash_{\mathbf{TS}} \Gamma \Rightarrow A$. Since $\Gamma \Rightarrow A$ was arbitrary, we may conclude that **TS** has no valid inferences or sequents. □

---

[8]By these rules we refer to the following, respectively.

$$\frac{\Gamma \Rightarrow \Delta}{\Gamma \Rightarrow \varphi, \Delta}\,[WR] \qquad \frac{\Gamma \Rightarrow \varphi, \Delta}{\Gamma, \neg\varphi \Rightarrow \Delta}\,[\neg L] \qquad \frac{\Gamma, \varphi \Rightarrow \Delta \quad \Gamma \Rightarrow \varphi, \Delta}{\Gamma \Rightarrow \Delta}\,[Cut^{\mathrm{A}}] \qquad \frac{\Gamma, \varphi \Rightarrow \Delta \quad \Sigma \Rightarrow \varphi, \Pi}{\Gamma, \Sigma \Rightarrow \Delta, \Pi}\,[Cut^{\mathrm{M}}]$$

[9]Whence, the aforementioned equivalence is witnessed e.g. by the following derivation, where the application of [*Cut*] is a rightful instance of both [Cut^{A}] and [Cut^{M}].

$$\dfrac{\dfrac{\emptyset \Rightarrow \neg A \qquad \dfrac{\emptyset \Rightarrow A}{\neg A \Rightarrow \emptyset}\,[\neg L]}{\emptyset \Rightarrow \emptyset}\,[Cut]}{\emptyset \Rightarrow B}\,[WR]$$

Fact 3.8 ([16, 26]) **ST** *and* **CL** *have the same set of valid inferences or sequents.*

About these logics we shall mention, in addition to the previous remarks, that in [4, 11, 25] it is shown that—through some suitable translation—the set of valid inferences in **LP** coincides with the set of valid meta-inferences in **ST**, while in [14] it is conjectured that—again, through some suitable translation—the set of valid inferences in $\mathbf{K_3}$, i.e. Strong Kleene logic,[10] coincides with the set of valid meta-inferences in **TS**. As Francesco Paoli pointed out to us, this conjecture was shown to be true, in light of the results proved in [33].

4 Philosophical Reflections

The previous discussion dealt with classical logic and three systems which, in light of the previously proposed criterion, might be legitimately called *paraconsistent*.

Certainly, that classical logic is *not*, but **LP** *is* paraconsistent should not surprise anyone, since these are well-known facts. Nevertheless, given our proposal, the previous remarks allow to offer a new look at the these systems. In this regard, we will say that **LP**, as well as **CL** adopt a *uniform* policy with regard to paraconsistency. We mean with this that, just like **CL** is not paraconsistent at either the inferential or the meta-inferential level, **LP** is both paraconsistent at the inferential and the meta-inferential level.

These remarks about uniformity suggest that it is reasonable to ask whether or not it is possible to have logics which have a non-uniform policy towards paraconsistency. A positive answer to this question has been offered in the previous sections. Two examples of the meaningfulness of this alternative are the substructural logics **TS** and **ST**. The former *is* paraconsistent, although it is not uniformly so, for it is paraconsistent at the inferential level, but not at the meta-inferential level. The latter *is* paraconsistent, although it is also not uniformly so, for it is paraconsistent at the meta-inferential level, but not at the inferential level.

Let us now comment on two philosophically discussions where the above remarks can have some interesting repercussions. Claiming that there are some paraconsistent logics which give a uniform and other that have a non-uniform policy with regard to the validity of Explosion is relevant to the discussion of *logical pluralism*: different levels of logical consequence can give different answers about the validity of a certain inference, rule, or scheme—in the case that concerns us, about Explosion. But, of course, these remarks can be generalized. As Barrio, Rosenblatt and Tajer [4] have shown, meta-inferential validity in **ST** coincides (through some suitable translation) with inferential validity in **LP**. If we also take into account that Cobreros, Ripley, Egré and van Rooij proved that **ST** and **CL** have the same set of valid inferences or sequents, this result can be interpreted conceptually as the admission that two *rival*

[10] That is, the matrix logic induced by the structure $\mathscr{M}_{\mathbf{K_3}} = \langle\{\mathbf{t}, \mathbf{i}, \mathbf{f}\}, \{\mathbf{t}\}, \{f^{\neg}_{\mathbf{K}}, f^{\wedge}_{\mathbf{K}}, f^{\vee}_{\mathbf{K}}\}\rangle$, such that $\langle\{\mathbf{t}, \mathbf{i}, \mathbf{f}\}, \{f^{\neg}_{\mathbf{K}}, f^{\wedge}_{\mathbf{K}}, f^{\vee}_{\mathbf{K}}\}\rangle$ is the 3-element Kleene algebra.

logics are *both right*, i.e. **CL** could be a correct response at the level of inferences and **LP** could be a right answer at the level of meta-inferences.

Obviously, we are *not* claiming that cases in which the inferential and the meta-inferential notion of validity come apart are a *proof* of logic pluralism. It is also possible to support the view according to which meta-inferential validity imposes conditions on inferential validity, as e.g. Dicher and Paoli [11] seem to maintain. Or that meta-inferential validity has no weight, for the only thing that matters is inferential validity logic, as Ripley and the defenders of **ST** seem to maintain. Instead, we affirm that logics like **ST** and **TS** open the possibility of adopting a *pluralistic* attitude about logic: depending on what level we are interested, different appropriate answers could be given.

Another important issue that the present discussion might have consequences for, is the question about the *meaning of logical connectives*, and—most importantly—the relation that their meanings have with their behavior in various inferential levels. For example, in **ST** modus ponens is valid for the conditional at the inferential level, but it is *not* a valid rule at the meta-inferential level, as Zardini [35] points out. If the meaning of a logical connective is given by the valid inferences in which it is involved, logics as **ST** seem to admit connectives with different meanings at different inferential levels.

Moreover, for *inferentialists* the question arises as to whether or not the meta-inferential properties of the logics (at least partly) determine the meaning of the connectives of the given system, as Dicher [10] seems to suggest. That is, if we compare the connectives of, for example, **TS**, **ST** and **CL** proof-theoretically[11]—following the remarks of e.g. [14]—do they have the same meaning, given they are equipped with the same set of operational rules? All these questions are of deep philosophical import, and we hope to discuss them in future work.

4.1 Answers to Some Possible Objections

To conclude this section, let us evaluate a number of objections that might be raised against our approach. We consider, initially, two objections which question that **TS** and **ST** are genuine paraconsistent logics. After that, we consider an objection that questions the extent to which our proposed criterion of paraconsistency is reasonable.

The first objection aims at **TS**, and it concerns whether or not it is a paraconsistent logic in a trivial sense. Everyone would accept (even if they do not accept our proposed characterization of a paraconsistent logic) that an inferential consequence relation with no valid inferences is paraconsistent. For Explosion is a (schematic) inference, and if no (schematic) inference is valid, a fortiori Explosion will be invalid for that logical consequence relation. This is, in fact, the situation with inferential validity in **TS**.

---

[11] Along these lines, **CL** can be proof-theoretically understood as Gentzen's sequent calculi **LK**, **TS** as **LK** minus the structural rule of Reflexivity, and **ST** as **LK** minus the structural rule of Cut.

Now, the objection goes: in which sense is a consequence relation with no valid inferences a *genuine* consequence relation? It seems that it is as meaningless as a consequence relation as the empty set itself, and—as Chemla, Egré and Spector suggest in [7]—it will be definitely non-standard to call the empty set a genuine consequence relation, let alone a *logic*. Furthermore, if a consequence relation with no valid inferences is not a genuine consequence relation, it hardly can represent a *genuine* paraconsistent consequence relation.

To this we reply by noticing, as is done in [7], that the fact that **TS** has no valid inferences does not allow to identify its inferential consequence relation with the empty set. For, in a certain sense, the fact that **TS** has no valid inferences is dependent on the language being employed. Were we to have a constant $\top$ representing the value **t**, and a constant $\bot$ representing the value **f**, then e.g. the following inferences will be valid in the referred extension of **TS** (and, thus, of the 3-valued Kleene algebra)

$$\top \Rightarrow \top \qquad \bot \Rightarrow \bot \qquad \bot \Rightarrow \top$$

More importantly, the addition of such constants to **TS** will *not* imply the validity of Explosion at the inferential level. Therefore, **TS** is a paraconsistent logic in a meaningful and non-trivial sense.

The second objection aims at **ST**, whose peculiarly non-uniform way of being paraconsistent has been called into question in some recent papers like [4, 11], causing an *impasse* regarding the qualification **ST** deserves as a classical or non-classical logic.

On the one hand, some of its advocates (i.e. Cobreros, Ripley, Egré and van Rooij) seem to claim that, given **ST** coincides with **CL** at the inferential level, then **ST** deserves to be referred *nothing more* than an alternative presentation of classical logic. This argumentative line appears to be supported by the fact that J.-Y. Girard employed in [16] the **ST** 3-valued q-matrix to give a presentation of classical logic where Cut fails.

On the other hand, some of its critiques, like Barrio, Rosenblatt and Tajer in [4] and Dicher and Paoli in [11] appear to think that **ST** should *not* be identified with classical logic, but with **LP**. This argumentative line appears to be supported also by the fact that e.g. Cobreros, Ripley, Egré and van Rooij usually present **ST** as Gentzen's sequent calculus **LK** for classical logic, *minus* the structural Cut rule. But if they prefer to talk about **ST** as a sequent calculus system, then they are prone to the following imputation due to Dicher and Paoli

> Notice, however, that in a sequent calculus *all of the action* takes place at the level of sequent-to-sequent rules, whereby from one or more sequents (intuitively understood as 'inferences') we derive more sequents (i.e., more 'inferences'). Which is to say, *the action takes place at the level of metainferences*. [11, p. 8, our emphasis]

The result of the previous dialectic is, then, that some say that **ST** is *not* paraconsistent, but is classical, because the only thing that matters is inferential validity and at that level **ST** coincides with **CL**, whereas some others say that **ST** *is* paraconsistent, but is not classical, because the only thing that matters is meta-inferential validity and at that level **ST** coincides with **LP**.

We stand in the middle: we take that both inferential and meta-inferential validity matter. Since both matter, then particularly meta-inferential validity matters and thus we think that **ST** deserves to be taken as a genuine paraconsistent logic. But do we draw a symmetric conclusion and claim that **ST** is classical? No. This might bother some objectors, and to the consideration of their potential objection we now turn.

Thus, the third objection concerns the extent to which our proposed criterion of paraconsistency is reasonable, and would run roughly as follows. It is unreasonable to say that a logic is paraconsistent if either its inferential or its meta-inferential consequence is, because this is an instance of a more general criterion that we would *not* accept, for it has instances that we would reject. Namely, the general criterion that

A LOGIC *is X if either its inferential or its meta-inferential consequence is X*

Now, the objection may continue, if *X* is 'classical', then we have just said that it would not be reasonable to accept that a logic is classical if e.g. its inferential consequence is *not* classical, but its meta-inferential consequence *is* classical.

To this we reply as follows. First of all, by claiming that a logic is paraconsistent if either its inferential or its meta-inferential consequence is we are *not* necessarily committed to accept the general criterion that a logic is *X* if either its inferential or its meta-inferential consequence is *X*. This is so, just like accepting an instance of the Law of Excluded Middle (e.g. 'Either Goldbach's Conjecture is true, or Goldbach's Conjecture is false') does *not* necessarily commit oneself to the unrestricted acceptance of the Law of Excluded Middle, for one may think that there are cases in which it may fail to hold (e.g. future contingents, etc.).

Finally, we do in fact think that there is a reason to refrain from adopting the general criterion, i.e. that some of its instances are wrong, in particular, the instance where *X* is 'classical'. We are of the opinion that a logic being classical at some inferential level *does not* propagate to a qualification of the entire logic, whereas a logic being non-classical—and, in particular, *paraconsistent*—*does* propagate to a qualification of the entire logic. The asymmetry resides, mainly, in the fact that being classical is a characteristic that requires the fulfillment of certain inferential features, while being non-classical and in particular paraconsistent is a characteristic that requires the *non*-fulfillment of certain inferential features. As is stressed by Ripley—in the quote of his that we mentioned in Sect. 1—paraconsistency is a *non*entailment claim, whereas it appears that classicality is an entailment claim. For this reason, it is reasonable for us to say that there is a difference in being paraconsistent, which requires that at least at *some* level (either the inferential or the meta-inferential) this nonentailment claim holds, and being classical, which seems to require that at *all* levels the entailment claim holds.

5 Conclusion

In the present paper we presented a new criterion for a logic to be *paraconsistent*: if either the inferential or the meta-inferential formulation of Explosion is invalid in **L**, then it is paraconsistent. Interestingly, we showed that a logic may invalidate one but validate the other, contrary to what happens in logics that have a uniform policy towards these matters, such as classical logic and **LP** which, respectively, validate both and validate neither of the formulations. The study cases that we focused on were two substructural logics, **TS** and **ST**; the former invalidates the inferential, but invalidates the meta-inferential formulation of Explosion, and the latter validates the inferential, but invalidates the meta-inferential version of Explosion. This strongly suggests that the proposed criterion is non-trivial and that there are interesting cases of logics which deserve to be called paraconsistent and that have not been regarded as such by the received view about paraconsistency, which focused exclusively on the inferential formulation(s) of Explosion. By focusing on versions of Explosion which are not inferential, but meta-inferential, we argued that Explosion comes in very different flavors and that it should be explored with greater generality that it has been, until now.

Let us close these conclusions with one final comment. In this paper we dealt with logical consequence between *formulae* and between *sequents*, thereby considering and evaluating inferential and meta-inferential versions of Explosion. But nothing prevents us from taking the investigation one step further and considering consequence relations between e.g. *meta-sequents*. Yet, again, if this is plausible, why stop there? We can definitely consider consequence relations between *meta-meta-sequents*, and so on and so forth. It can be easily seen how this procedure can be further reproduced, giving us a whole hierarchy of inferences concerned with the logical relations between objects of the lower level(s). In doing so it is interesting to, thus, look at inferences as having, or being of, some level represented by some ordinal number. Common inferences relating formulae are, therefore, of level 0, whereas meta-inferences are of level 1, meta-meta-inferences are of level 2, and so on and so forth. In this vein, Explosion might be regarded as a meta-schematic inference $A, \neg A \Rightarrow^{\alpha} B$, for α an ordinal. In other words, as a meta-scheme or scheme of schemes, i.e. a scheme that gives, for each ordinal, a schematic inference, namely the formulation of Explosion for that inferential level.

These surely are interesting directions to explore. A full exposition of them will require defining how big the hierarchy is and if it has a fixed point or not, how inferences at some peculiar levels (e.g. at limit ordinals) look, how do the formulations of Explosion beyond the meta-inferential level look, and many other technical and conceptual matters. Settling this issues is no doubt an interesting task, but one which demands an amount of space beyond the one available for this paper. We hope to investigate them in further research.

Acknowledgements The ideas contained in this paper were presented at the Work In Progress seminar of the Buenos Aires Logic Group, the XVI Trends in Logic conference, and the Kyoto Philosophical Logic Workshop II at Kyoto University. We are thankful to the members of those audiences for the discussion provided. We would also like to thank Rohan French, Andreas Fjellstad, Hitoshi Omori, Francesco Paoli, Graham Priest, Dave Ripley, Zach Weber and two anonymous reviewers for their helpful suggestions and comments on previous versions of this article. Finally, we are thankful to the editors of this volume, Walter Carnielli, and Jacek Malinowski, for their assistance through the editorial process.

References

1. Asenjo, Florencio. 1966. A calculus of antinomies. *Notre Dame Journal of Formal Logic* 7 (1): 103–105.
2. Avron, Arnon. 1988. The semantics and proof theory of linear logic. *Theoretical Computer Science* 57 (2–3): 161–184.
3. Avron, Arnon. 1991. Simple consequence relations. *Information and Computation* 92 (1): 105–139.
4. Barrio, Eduardo, Lucas Rosenblatt, and Diego Tajer. 2015. The logics of strict-tolerant logic. *Journal of Philosophical Logic* 44 (5): 551–571.
5. Blok, Willem, and Bjarni Jónsson. 2006. Equivalence of consequence operations. *Studia Logica* 83 (1): 91–110.
6. Carnielli, Walter, and Marcelo Coniglio. 2016. *Paraconsistent Logic: Consistency, Contradiction and Negation*. Dordrecht: Springer.
7. Chemla, Emmanuel, Paul Egré, and Benjamin Spector. 2017. Characterizing logical consequence in many-valued logics. *Journal of Logic and Computation* 27 (7): 2193–2226.
8. Cobreros, Pablo, Paul Egré, David Ripley, and Robert van Rooij. 2012. Tolerant, classical, strict. *Journal of Philosophical Logic* 41 (2): 347–385.
9. Cobreros, Pablo, Paul Egré, David Ripley, and Robert van Rooij. 2014. Reaching transparent truth. *Mind* 122 (488): 841–866.
10. Dicher, Bogdan. 2016. A proof-theoretic defence of meaning-invariant logical pluralism. *Mind* 125 (499): 727–757.
11. Dicher, Bogdan, and Francesco Paoli. 2017. ST, LP, and Tolerant Metainferences. (Unpublished Manuscript).
12. Frankowski, Szymon. 2004. Formalization of a plausible inference. *Bulletin of the Section of Logic* 33 (1): 41–52.
13. Frankowski, Szymon. 2004. p-consequence versus q-consequence operations. *Bulletin of the Section of Logic* 33 (4): 197–207.
14. French, Rohan. 2016. Structural reflexivity and the paradoxes of self-reference. *Ergo* 3 (5): 113–131.
15. Gabbay, Dov. 1996. *Labelled Deductive Systems*. Oxford: Oxford University Press.
16. Girard, Jean-Yves. 1987. *Proof Theory and Logical Complexity*. Napoli: Bibliopolis.
17. Hansson, Sven Ove. 2016. Logic of belief revision. In *The Stanford Encyclopedia of Philosophy*, winter 2016 ed, ed. Edward N Zalta. Stanford University. http://plato.stanford.edu/archives/win2016/entries/logic-belief-revision/.
18. Humberstone, Lloyd. 2011. *The Connectives*. Cambridge: MIT Press.
19. Malinowski, Grzegorz. 1990. Q-consequence operation. *Reports on Mathematical Logic* 24 (1): 49–59.
20. Malinowski, Grzegorz. 2007. THAT $p + q = c$(onsequence). *Bulletin of the Section of Logic* 36 (1/2): 7–19.
21. Malinowski, Grzegorz. 2014. Kleene logic and inference. *Bulletin of the Section of Logic* 43 (1/2): 43–52.
22. Mares, Edwin, and Francesco Paoli. 2014. Logical consequence and the paradoxes. *Journal of Philosophical Logic* 43 (2–3): 439–469.
23. Priest, Graham. 1979. The logic of paradox. *Journal of Philosophical logic* 8 (1): 219–241.

24. Priest, Graham, Koji Tanaka, and Zach Weber. 2016. Paraconsistent logic. In *The Stanford Encyclopedia of Philosophy*, Winter 2016 ed, ed. Edward Zalta. Stanford University. http://plato.stanford.edu/archives/win2016/entries/logic-paraconsistent/.
25. Pynko, Alexej. 2010. Gentzens cut-free calculus versus the logic of paradox. *Bulletin of the Section of Logic* 39 (1/2): 35–42.
26. Ripley, David. 2012. Conservatively extending classical logic with transparent truth. *The Review of Symbolic Logic* 5 (02): 354–378.
27. Ripley, David. 2013. Paradoxes and failures of cut. *Australasian Journal of Philosophy* 91 (1): 139–164.
28. Ripley, David. 2015. Paraconsistent logic. *Journal of Philosophical Logic* 44 (6): 771–780.
29. Scott, Dana. 1971. On engendering an illusion of understanding. *The Journal of Philosophy* 68 (21): 787–807.
30. Shoesmith, David, and Timothy Smiley. 1978. *Multiple-Conclusion Logic*. Cambridge: Cambridge University Press.
31. Shramko, Yaroslav, and Heinrich Wansing. 2010. Truth values. In *The Stanford Encyclopedia of Philosophy*, Summer 2010 ed, ed. Edward Zalta. Stanford Uniersity. http://plato.stanford.edu/archives/sum2010/entries/truth-values/.
32. Shramko, Yaroslav, and Heinrich Wansing. 2011. *Truth and Falsehood: An Inquiry into Generalized Logical Values*. Dordrecht: Springer.
33. Stärk, Robert. 1991. A complete axiomatization of the three valued completion of logic programs. *Journal of Logic and Computation* 1 (6): 811–834.
34. Urbas, Igor. 1990. Paraconsistency. *Studies in Soviet Thought* 39 (3/4): 343–354.
35. Zardini, Elia. 2013. Naive modus ponens. *Journal of Philosophical Logic* 42 (4): 575–593.

Provided You're not Trivial: Adding Defaults and Paraconsistency to a Formal Model of Explanation

David Gaytán, Itala M. Loffredo D'Ottaviano and Raymundo Morado

Abstract Let us assume that a set of sentences explains a phenomenon within a system of beliefs and rules. Such rules and beliefs may vary and this could have as a collateral effect that different sets of sentences may become explanations relative to the new system, while other ones no longer count as such. In this paper we offer a general formal framework to study this phenomenon. We also give examples of such variations as we replace rules of classical deductive logic with rules more in the spirit of da Costa's paraconsistent calculi, Reiter's default theories, or even a combination of them. This paper generalizes the notion of epistemic system in [6]. That notion was used to analyze the concept of explanation, using Reiter's default theories and a specific paraconsistent logic of da Costa. Our proposal is a formal framework, **GMD**, based on doxastic systems, which allows us to analyze the interaction between theoretical constructs (in this case, explanations), theories and logics. We mention some obstacles, we develop the formal framework, and finally we apply it to the modeling of scientific explanation. Along the way, we try to shed light on different kinds of interaction between paraconsistency and non-monotonicity.

1 Our Roadmap

In this paper we present a formal framework motivated by the problems of modeling explanation from the point of view of Philosophy of Science. The original aim was to help deal with some counterexamples to the classical models of explanation of Hempel and Oppenheim and handle the difficulties that some theorists observe in

D. Gaytán (✉)
Universidad Autónoma de la Ciudad de México, Mexico City, Mexico
e-mail: david.gaytan@uacm.edu.mx

I. M. L. D'Ottaviano
Universidade Estadual de Campinas, Campinas, Brazil
e-mail: itala@cle.unicamp.br

R. Morado
Universidad Nacional Autónoma de México, Mexico City, Mexico
e-mail: morado@unam.mx

W. Carnielli and J. Malinowski (eds.), *Contradictions, from Consistency to Inconsistency*, Trends in Logic 47, https://doi.org/10.1007/978-3-319-98797-2_6

their analysis. We generalize previous results with respect to explanation and develop a more general formal framework that could prove useful in analyzing other scientific constructs like scientific laws, predictions, or general techniques to gather evidence. This formal framework can be fruitful to analyze the relationship between explanations and inconsistent theoretical contexts. In particular, its modular presentation allows our formal framework to capture relations between explanation and context. Context is represented as composed of elements whose modulation in the formal framework plays a very important role.

The formal framework is based on the logical interaction between a non-mononotonic logic and a paraconsistent logic. In particular, both systems used are amongst the first ones proposed. In the case of paraconsistent logic we will use the hierarchy of calculi of Newton da Costa [4]. In the case of non-monotonic logic, we will use the logic of default reasoning of Raymond Reiter [17]. But it is important to note that the formal framework presented in this work, with tiny modifications, could be put in a logical interaction with other non-monotonic and paraconsistent systems.

An obstacle to the logical interaction between a non-monotonic logic and a paraconsistent logic is that several non-monotonic logics require consistency to reach a conclusion. To address this problem, we include a general rule structure based on Reiter's defaults. We then build the notion of a doxastic system as a formal environment for the interaction between logics and theories. We assume theories have underlying logics and that we can distinguish between logics and theories. Next, we show the possibility of having non-monotonic doxastic systems, and paraconsistent non-monotonic doxastic systems. Finally, we apply this to the design of a paraconsistent non-monotonic formal model of explanation.

2 Example of Some Difficulties in Modeling Explanations

Scientific explanation is a phenomenon studied by, among other disciplines, Artificial Intelligence and Philosophy of Science. Both approaches can be viewed as complementary. In the context of Artificial Intelligence, explanations have been seen mostly as a process to infer a hypothesis. In Philosophy of Science, explanations have been assumed to be a kind of product of scientific activity.[1] In this tradition, the central idea is to characterize explanation in Science through either a theory or a formal model. When the choice is to construct a model, the classical proposal is to regard explanations as argumentative structures. In this approach the central issue is: What idea of argumentation can clarify the notion of explanation? Can we build a model of argumentation able to characterize an explanation adequately? The model is expected to help us to identify explanations and to distinguish them from other things in Science. This line of research in Philosophy is expected to help outline a notion of scientific rationality. In [7] are proposed the "classical models", basically

[1] Classical texts are [7, 8].

a pair of argumentative schemas that, adding other relevant conditions, are viewed as an abstract representation of scientific explanations. These schemas contain laws and initial conditions as premises, and a description of a phenomenon as a conclusion. This kind of argument form, adding non-vacuity conditions and empirical restrictions, yields the classical models. The set of premises is called the "*explanans*" and the conclusion is called the "*explanandum*". Thus, the argumentative analysis of explanation proposes a structure of the following kind:

explanans, therefore, *explanandum*.

The two basic classical models use either deductive or statistical laws in the *explanans*. If the laws are deductive, we have a "deductive nomological model" (D-N Model). On the other hand, if the laws have a statistical character, we have an "inductive statistical model" (I-S Model).[2]

Parallel to these developments, some philosophers advanced their own formal models of explanation.[3] Some of these models were argumentative models, but some dispensed with the assumption that an explanation can be modeled as an argument.[4] The modeling of explanation has run into roughly three kinds of problems[5]:

(I) Problems about the inferential relation in the model.[6]
(II) Problems about the relation of relevance between *explanans* and *explanandum*.[7]
(III) Problems about the context of explanation.[8]

Types I and III are problems closely linked to non-monotonicity. Some classical counterexamples presented against the classical models of Hempel, and Hempel's own analysis about the problem of epistemic ambiguity in [8], implicitly pointed out the convenience of relativizing explanations to knowledge bases under change. A first example is [16], which tries to characterize probabilistic and random explanations. Despite attempting a deductive modeling, the author recognizes the difficulties of the task. In [20] the necessity of a non-deductive model is even clearer. Van Fraassen supports the characterization of explanations as counterfactuals, proposed by David Lewis in [10], as a viable position in principle, and also characterizes the contextual dependence of explanations via a set of more changeable assumptions and a set of less changeable ones. One of the consequences of his analysis is that we need to explore conditionals in which the reinforcement of the antecedent is not valid. Finally, in

[2]These models may also differ in other aspects.

[3]E.g. Salmon, Railton, and van Fraassen. For a good historical survey of these philosophical discussions, see [18].

[4]For example, van Fraassen, [20].

[5]For a reconstruction of classical problems about scientific explanation, see [6].

[6]We are thinking about the counterexamples about sufficient or necessary characteristics of the model, or about the so called "epistemic ambiguity problem" (Cf. [8]), and also about the hard problems of deductive representation in Science, reported independently in [9] and [15].

[7]E.g., the problems about causal underdetermination or about asymmetries of explanation. For both topics you can see [1].

[8]For the notion of causality in different contexts, see [20].

[19], this same intuition is analyzed. However, in none of these cases a decidedly non-monotonic modeling, which would take into account the problems indicated in the literature, was attempted.

In order to show the kind of problems we are considering, we use a common example of explanation from Philosophy of Science's literature:

Example **EX**

(P1) John ate arsenic two days ago.

(P2) Anyone who eats arsenic will die in a lapse of 24 hours.

(C) John died

The example fulfills the conditions of an explanation in Hempel's model. The *explanans* is the set P1, P2 of premises, and we can consider that the *explanandum* is the conclusion C of the argument, and we have a deductive or highly probable relation of derivation between that set and C. Hempel also included into his model several conditions of adequacy. One of them is

(AC1) each member of the *explanans* must be true.

Taking into account some counterexamples to Hempel's models of explanation, we can demand further adequacy conditions. The following three are candidates to adequacy conditions that go beyond Hempel's original requirements, although they do not seem to be necessary for every kind of scientific explanation.

(AC2) the *explanans* has to give a causal account of the *explanandum*.

(AC3) the *explanans* must describe facts that actually are the causes of the occurrence of the *explanandum*.

(AC4) the *explanans* must be related to the *explanandum* by means of at least a high probability derivation.

One may have at least two different attitudes towards AC1–AC4. One may consider them as necessary conditions if an argument is to be an explanation at all. Another possible perspective is to consider that AC1–AC4 constitute conditions for some argument to be a *good* explanation. We shall adopt this latter perspective.

(PS1) Imagine that Joseph wanted to play a joke on John: Joseph has made John believe some P1 (for instance, that fire is produced by phlogiston, or that light travels in space through an ether), and actually P1 is false. In this situation, erroneously, John believes that P1 is true, i.e., John believes that AC1 is fulfilled. Hempel would discard the case as a pseudo-explanation since P1 is false, but we have a strong intuition that those historic explanations were erroneous yet explanations nonetheless. We do not necessarily want to consider P1 as a true proposition in order to consider Ex1 an explanation. Maybe we do not need to consider every premise to be true in order for an argument to qualify as an explanation.

(PS2) Let us now suppose that P1 and P2 are true. Note, also, that P1 and P2 give a causal justification for C. Then, example EX counts as an explanation; even more, as a true and causal explanation. We have an argument that fulfills the adequacy conditions AC1 and AC2. Nevertheless, there are a lot of possible meanings of "causal relation" and we may consider also some relations fulfilled in an explanation that are not causal relations. We have then different possibilities for what an explanatory relevance relation needs to be.

(PS3) Imagine that Example EX fulfills AC1–AC3, but Joseph knew that John had cardiac problems. AC3 is usually considered a condition to prevent this possible failure. Then, when Joseph is informed that his friend is dead, he formulates an explanation that appeals not to arsenic but to John's suffering from heart problems. Would we say that the argument of Joseph is not an explanation? Similarly to the PS1 case, we have a strong intuition that those historic explanations were erroneous or bad explanations but yet explanations nonetheless. In addition, if we maintain that explanations function on epistemic level, in the sense that explanations are doxastic proposals to account for a particular fact, Joseph had an explanation. Maybe Joseph's explanation does not correspond to reality, but it could be considered a correct formulation of an explanation. May be an argument representing an explanation does not need to refer to a cause chain that actually causes the *explanandum*.
(PS4) Let us, finally suppose that P2 in Example EX is rather "Anyone that eats arsenic will likely die in a lapse of 24 hours". We consider also that Example EX, with the new P2, fulfills AC1–AC3. But, let us also suppose that the quantity of arsenic consumed by John, made the probability in P2 a low probability. Maybe we know, in this hypothetical situation, that such amount of arsenic has a little effect in very few people. In a certain sense, the fact that John consumed arsenic explains his death, but there is a possibility that Example EX, although it is a good formulation of an explanation, may be a bad explanation of John's death. Example EX is an explanation but maybe not a good one. In this work, we will consider that even in such situation, Example EX is an explanation.

We may accept conditions of the type of AC1–AC4 as necessary conditions for the goodness of an explanation according to our beliefs and rules.[9] Assuming that the goodness of an explanation depends on the beliefs and rules from which such goodness is evaluated, the doxastic context of an explanation, containing as it does such beliefs and rules, might generate different adequacy conditions. In order to distinguish different sets of conditions for a good explanation it is very important to take into account the variations that the adequacy conditions might undergo. These variations might change what it means to be a good explanation. Because of this a modulating treatment might prove useful. The way we propose to handle explanations will include representing several internal elements of a minimal explanation and other context related components so that the elements are subject to variance, to be "modulated". Because our proposal allows for such variation, our treatment modulates and not just models explanations. As we shall also see, the variation of the components must be reflected in changes in the form of interaction between some of the components; therefore, our model will capture a modulation between components and not only of them in isolation.

[9]PS1–PS4 are taking from the discussion in Philosophy of Science.

3 Building Rules and Doxastic Systems on GMD

An integral part of our paper is building simple examples of modulation. For that we will use a simple construct that will reflect a system of beliefs, namely a doxastic system, with a dynamic internal structure. A doxastic system contains both belief sets and methods to change those sets. We can capture those methods by means of the idea of a rule. We will concentrate on the representation of modulations in the dynamics of rules and its consequences inside the doxastic system as a whole.

We intuitively understand here the general structure of a rule in a doxastic system S, as a triple $\langle A, \gamma, M\rangle$, such that A is a set of formulas sufficient to obtain γ, γ is the formula obtained normally from A, and M is the modulation of the rule. A modulation in the rule is understood as a condition in the rule that allows or restricts the inference of γ from A. This M may be as variegated as societal constraints, resource limitations, ethical considerations, etc. In this paper we will consider only simple modulations such as conditions of consistency or conditions of non triviality.

We now construct a formal framework to relate these intuitive concepts of a doxastic system and of a rule. We will call this framework "**GMD**" after the last names of the authors.

We begin assuming, as a base to **GMD**, a standard first order logical vocabulary: logical constants $V_1 = \{\rightarrow, \wedge, \vee, \leftrightarrow, \neg, \forall, \exists\}$, constants for individuals $V_2 = \{a, b, \ldots, t, a_1, \ldots, t_1, a_2, \ldots\}$, variables for individuals $V_3 = \{u, \ldots, z, u_1, \ldots, z_1, u_2, \ldots\}$, constants for predicates $V_4 = \{P, Q, R, \ldots, Z, P_1, \ldots, Z_1, P_2, \ldots\}$, and propositional variables $V_5 = \{\alpha, \beta, \ldots, \omega, \ldots, \alpha_1, \ldots, \omega_1, \alpha_2, \ldots\}$. Since our vocabulary is going to be $\bigcup_{i=1}^{5} V_i$, we construct the set of well formed formulas (wffs), FOR, as usual.

We will now propose more specific versions of those intuitive ideas. We will construct an abstract notion of rule. Also a formal environment for the interaction between logics and theories: the doxastic systems.

Definition 1 (*Definition* 1) An (specific) intra-system rule ("*IR*") for **GMD**, *IR*, is an ordered triple $\langle A, \gamma, B\rangle$ where:
(1) $A, B \subset FOR$,
(2) A and B are finite,
(3) $\gamma \in FOR$,
(4) There exists a set $K \subseteq FOR$ and a relation (condition) C (of $\langle A, \gamma, B\rangle$) between B and K, such that:
If every element $\alpha \in A$ is obtained (or appears in a previous line in a given demonstration), and C is fulfilled between B and K, then γ is obtained (in this case, we say that γ is inferred from A by means of $\langle A, \gamma, B\rangle$). This C can be considered the modulation of $\langle A, \gamma, B\rangle$.

For a given *IR* $\langle A, \gamma, B\rangle$, we will call every $\alpha \in A$ "a prerequisite" of the *IR*, γ the "conclusion" of the *IR*, A the "presence condition" of the *IR*, every $\beta \in B$ a "proviso" in the *IR*, B the "absence condition" of the *IR*, $A \cup B$ the "requisites" of the *IR*, and K, a "base of provisos" of the *IR*.

Note that by means of our abstract notion of rule, based in a structure $\langle A, \gamma, B\rangle$, we can represent standard rules on FOR, as cases of rules with $B = \emptyset$. In such cases, we write $[A, \gamma]$ instead of $\langle A, \gamma, B\rangle$. (Otherwise it is considered a non-standard rule).

Remark 1 The general intuition in **Definition** 1 is that some rules have a structure similar to the classical one, in the sense that the conclusion depends only on a set of prerequisites or premises; and we have rules with a default structure, in which the conclusion could depend also on another kind of conditions, the provisos.

With this notion of rule we can represent and modify the mechanisms of inference in reasoning considered with possible restrictions. The modifications made inside the structure of rules could facilitate the representation of reasoning, and environments with contradictions and with the possibility of losing conclusions (what usually could be called "paraconsistent reasoning" and "non-monotonic reasoning"). We can think of these environments in terms of an interaction between theories and logics (what we will call in **GMD** "mixed doxastic systems"). In this line of thought, one of our aims is to represent theories containing contradictions (inconsistent theories), but also weak inferences (may be defaults) that could generate contradictions in our theoretical corpus, and even logically closed inconsistent sets of propositions, generated from initial theoretical and logical conditions (inconsistent extensions). We want to distinguish between "permitted" contradictions, and "not permitted" contradictions (which justify a paraconsistent formal environment in GMD). If a contradiction $\alpha \wedge \neg\alpha$ is permitted in an initial system of beliefs, then we say, closely to the use of da Costa, that α is a "bad behaved" formula in that system. On the contrary, if $\alpha \wedge \neg\alpha$ is not permitted in that initial system of beliefs, then we say that α is a "well behaved" formula in that system. Both last two terms are due to da Costa.

With modifications inside the rule (what we will call "modulations"), we want to represent weak (non deductive), non-monotonic and paraconsistent reasoning. We think that this kind of representation of reasoning can contribute to model a minimal argumentative notion of explanation. We will make all these modifications in a manner that allows us to capture combinations of theories and logics of different kinds, and also different kinds of interactions between theories and logics.

Definition 2 (*Definition* 2) An ordered pair $\langle \Psi, \Omega\rangle$ is a doxastic system (ds) Σ in **GMD** iff:

(1) $\Psi \neq \emptyset$ is a set of IR rules and $\Omega \subseteq FOR$. We may understand Ψ as the inferential engine and the elements of Ω as the sentences in Σ.

(2) Ψ generates, through applications of its elements, a consequence function Cn_Ψ such that its domain is FOR and its co-domain is $\wp\wp(FOR)$, i.e., $Cn_\Psi : FOR \longrightarrow \wp\wp(FOR)$. In particular, it is clear that $\Omega \in \wp(FOR)$ and that $Cn_\Psi(\Omega) \subseteq \wp(FOR)$. $Cn_\Psi(\Omega)$ is called the set of extensions of $\langle \Psi, \Omega\rangle$.

In a ds $\Sigma = \langle \Psi, \Omega\rangle$, Ψ may represent the underlying logic Λ of Σ. We will consider Λ as a standard underlying logic if every rule $IR \in \Psi$ is an standard rule on FOR. (Otherwise it is considered a non-standard underlying logic). By the way, if some rules IR in Ψ have an empty presence condition, $A = \emptyset$, and an empty absence condition, $B = \emptyset$, they function as axioms, and Λ may be considered an axiomatic logic.

In a doxastic system $\Sigma = \langle \Psi, \Omega \rangle$, Ω may represent a theory T, as a set of beliefs.

Definition 3 (*Definition* 3) A doxastic system $\Sigma = \langle \Lambda, T \rangle$, is called "a mixed doxastic system", mds, in **GMD** iff at least two rules, $\langle A_1, \gamma_1, B_1 \rangle_i, \langle A_2, \gamma_2, B_2 \rangle_j \in \Lambda$ are such that $B_1 = \emptyset$ and $B_2 \neq \emptyset$. In other words, a mds is a ds with at least one standard rule and at least one non-standard rule.

Definitions 2 and 3 are generalizations of the idea of epistemic system developed in [6], which was applied to a model scientific explanation. That work captured the interaction between the calculus C_1, from the hierarchy of calculi of Newton da Costa, with a logic of default reasoning of Raymond Reiter. These new definitions are made in order to capture interactions between different logics and different sets of beliefs (or theories), also in connection with the notion of explanation.

Definition 4 (*Definition* 4) A derivation of a wff ϕ from a set of formulas $Z \subseteq FOR$ in a mds $\Sigma = \langle \Lambda, T \rangle$ is a sequence of wffs $\langle \gamma_1, \ldots, \gamma_n \rangle$ such that:
(1) $\gamma_n = \phi$.
(2) For every γ_i in the sequence, at least one of the following conditions holds:
(a) $\gamma_i \in T \cup Z$.
(b) γ_i follows syntactically from $T \cup Z$ according to Λ by Cn_Λ of $\langle \Lambda, T \rangle$. We can also formulate this as saying that γ_i follows syntactically from Z by Cn_Σ.
We say that there is a $\langle \Lambda, T \rangle$ -derivation ϕ from Z, or that there is a Σ-derivation of ϕ from Z. In symbols, $Z \vdash_\Sigma \phi$.

The structure of the notion of derivability relative to a doxastic system allows us to represent several interesting situations. For $\Sigma = \langle \Lambda, T \rangle$, $Z \vdash_\Sigma \phi$ can represent the following cases[10]:

(1) With $Z = \emptyset$, represents the case in which ϕ follows from Σ,
(2) with $\Lambda = \emptyset$, the case in which ϕ is a reiteration, i.e., $\phi \in Z \cup T$ (a reiteration of the theory or the additional information).
(3) with $T = \emptyset$, the case in which ϕ follows from Z only by the logical resources of Σ,
(4) with $Z \cup T = \emptyset$, the case in which ϕ is a logical truth in Σ,
(5) with $Z \cup \Lambda = \emptyset$, the case in which ϕ is an assumption of Σ, i.e., $\phi \in T$ (an assumption of the theory T of Σ),
(6) with $\Lambda \cup T = \emptyset$, the case in which ϕ is an additional premise for Σ, i.e., $\phi \in Z$.

Remark 2 The idea of keeping Z explicitly out of the doxastic system is to be able to represent how different sets of propositions may interact with different doxastic

[10] As can be noted by the use we make of T, we can think of the theory T as the set of fundamental assumptions of the theory. These assumptions do not necessarily have to be identified with axioms. They could be the axioms of the theory, but they could be also some of its fundamental consequences: basically, we can understood T as the set of assumptions about the world.

systems. For instance, we may want to analyze if a certain set Z can constitute an *explanans* with respect to a particular theory T. In some cases, we may also want to analyze if a certain set Z constitutes an explanation regarding a particular theory T with a particular underlying logic Λ. In the analysis from **GMD**, the results of these combinations (*explanans* Z, theory T, underlying logic Λ) could be crucially different. The idea is that two important elements of some contexts are their logic and their presuppositions. We try to provide a highly simplified representation of these elements with the notion of doxastic system, and our notion of a Σ-derivation is a pale shadow of that of a contextual derivation.

In terms of consequence functions we have:
The set of extensions from Z relative to $\langle \Lambda, T\rangle$, denoted by $Cn_{\langle \Lambda,T\rangle}(Z)$, is defined as follows:
$Cn_{\langle \Lambda,T\rangle}(Z) = \{E \mid E \text{ is an extension of } \langle \Lambda, T \cup Z\rangle\}$.
For $\Sigma = \langle \Lambda, T\rangle$, we will write:
$Cn_{\Sigma}(Z) = \{E \mid E \text{ is an extension of } \langle \Lambda, T \cup Z\rangle\}$.

We can define the syntactic notion of consequence relative to different logics Λ and different theories T. Thus, our central strategy in order to compare explanations in theoretical contexts is: In order to represent a theory we can take T as something like a base belief set, and, to represent a logic, we can use Λ. In particular, we are interested in using as our Λ a combination of a default theory in the sense of Raymond Reiter and C_1 of the hierarchy of calculi $C_n^{*=}$, $1 \leq n < \omega$, of Newton da Costa.

A particular mds can contain rules of different types generated by the general IR structure of an ordered triple. We can define classical rules, with certain modulations in B, and we can define default rules with certain other modulations in B.

The general idea is to generate, as a particular case of mds, each default theory: "Reiter's mds". And we can also capture, with modifications in the way we apply the defaults, systems we have called "Reiter-da Costa's mds".

We can characterize different kinds of interaction between logics and theories depending upon the construction of the consequence relation. Because of this it is possible to generate standard default theories like Reiter's, or allow for interaction between non-monotonicity and paraconsistency. This shows, again, the possible modulations in our proposal, by means of the description of the function of consequences Cn mentioned above, not only in (i) our general structure of a rule, but also in (ii) the theories and logics. Let's show these modulations in more detail in the next two sections.

4 Building a Default Theory on GMD

In this section we use our formal framework **GMD**, based upon an abstract formulation of rules and upon the idea of the interaction between theories and its underlying logics, in order to represent standard default logic in Reiter's line of thought.

In first place, we assume First Order Classical Logic with identity and functions by means of the axiomatic structure of the system $C_0^{*=}$ from da Costa hierarchy $C_n, 0 \leq n \leq \omega$ [4]. This system corresponds to First Order Classical Logic. We will understand this system as a set of rules $[A, \gamma]$, some of them with $A = \emptyset$.

In second place, we present a natural generalization of default logic by means of $C_0^{*=}$ and a generalization of rules described above.

Now we construct a more specific yet still very general case of extension.

Definition 5 (*Definition* 5) Let $E, Z \subseteq FOR$, and the mds $\Sigma = \langle \Lambda, T \rangle$, $E_0 = T \cup Z$ and $E_{i+1} = Cn_{\langle \Lambda, \emptyset \rangle}(E_i)$, for all E_i, where $i \geq 0$; then we say that a set E is an extension of $\langle \Lambda, T \cup Z \rangle$ iff $E = \bigcup_{i=0}^{\infty} E_i$.[11]

For $S \subseteq FOR$, from now on we may write $Cn_\Lambda(S)$ instead of $Cn_{\langle \Lambda, \emptyset \rangle}(S)$.

Remark 3 **Definition** 5 does not restrict the kind of theory or underlying logic used in the mds. Although we will focus on the case of default and paraconsistent logics, this **GMD** framework is general enough to accommodate mixed doxastic systems based on other kinds of Λ. Note also the very general idea of the syntactic consequence relation between E_i and E in $Cn_\Lambda(E_i)$. This relationship will be clearer when we specify it in the case of default reasoning.

We give now the usual notions related to consistency. Given a set $S \subseteq FOR$, we call S a "consistent set of formulas" iff it does not exist any wff θ such that $\theta \in \bigcup Cn_\Lambda(S)$ and also $\neg\theta \in \bigcup Cn_\Lambda(S)$, where Λ is classical logic. Otherwise, we call S an inconsistent set of formulas.

The set of all formulas, *FOR*, is called "the trivial set". Obviously, If S is the trivial set, then S is inconsistent.

Given an mds Σ, we call Σ an inconsistent mds iff every extension E of Σ is inconsistent; if it is not the case, we call it a non-inconsistent mds.

Given an mds $\Sigma = \langle \Lambda, T \rangle$, we call Σ a "trivial mds" iff every extension E of Σ is trivial (i.e. every $E = FOR$); if this is not the case, we call it a "non-trivial mds".

Given an mds Σ, we call Σ an "explosive mds" iff for every inconsistent set S, every $E \in Cn_\Sigma(S)$ is trivial (i.e. every $E = FOR$).[12] If it is not the case, we call it a "paraconsistent mds".

Theorem 1 *Given* $\Sigma = \langle \Lambda, T \rangle$ *on* **GMD** *and* $Z \subseteq FOR$, *and the following conditions:*
(i) $\Lambda = C_0^{*=}$,
(ii) every $IR \in \Lambda$ *is a standard rule (i.e., with* $B = \emptyset$*),*
then not necessarily every extension $E \in Cn_\Sigma(Z)$ *is a consistent extension.*
Sketch of proof:
Let $T = \{\alpha, \neg\alpha\}$. *Then, every extension* $E \in Cn_\Sigma(Z)$ *is such that* $\{\alpha, \neg\alpha\} \subseteq E$. *Therefore, E is inconsistent.*

---

[11] Note the emphasis: The only sets that are considered extensions are fixed points of the consequence function.

[12] Please note that we are talking here of a trivial mds and not of a trivial Cn_Σ function.

Remark 4 It is important to note that, in our formal framework **GMD**, not every sr rule is necessarily a classical rule, and also that to have standard rules does not guarantee that a system will produce only consistent sets of consequences.

Theorem 2 *In the initial conditions of the* **Theorem** *1 above, $\langle \Lambda, T \cup Z\rangle$ has an inconsistent extension E iff E = FOR.*
Sketch of proof:
$\Rightarrow$
Let $T = \{\alpha, \neg\alpha\}$. Then, by **Theorem** *1, every extension $E \in Cn_\Sigma(Z)$ is inconsistent. Given i, for every $E \in Cn_\Sigma(Z)$, and $\beta \in FOR$, $(\alpha \wedge \neg\alpha) \rightarrow \beta \in E$, therefore, $\beta \in E$.*
$\Leftarrow$
Let an arbitrary E, E = FOR. Then $\{\alpha, \neg\alpha\} \subseteq E$. Then E is inconsistent.

Theorem 3 *In the initial conditions of the* **Theorem** *1 above, $\langle \Lambda, T \cup Z\rangle$ has an inconsistent extension E iff $T \cup Z$ is inconsistent.*
Sketch of proof:
$\Rightarrow$
Let E^1 be inconsistent, then there exists $\alpha \in FOR$, such that $\{\alpha, \neg\alpha\} \subseteq E^1$. Now assume $T \cup Z$ is consistent. Given i and ii, Λ is classical logic and its rules are sr. Then, given our assumption, either α or $\neg\alpha$ is derived by rules in Λ. Therefore, there exists a valid argument that shows $T \cup Z \vdash_\Sigma \alpha \wedge \neg\alpha$. But this is a contradiction, because Λ is classical logic and its rules are sr, and in these conditions $T \cup Z \vdash_\Sigma \alpha \wedge \neg\alpha$ is not a valid argument. Therefore, $T \cup Z$ is inconsistent.
$\Leftarrow$
Since every extension $E \in Cn_\Sigma(Z)$ is such that $T \cup Z \subseteq E$, therefore, if $T \cup Z$ is inconsistent then E is inconsistent.

Theorem 4 *In the initial conditions of* **Theorem** *1,*
$\Sigma = \langle \Lambda, T \cup Z\rangle$ has a trivial extension E iff $T \cup Z$ is inconsistent. (E is the only extension.)
Sketch of proof:
$\Rightarrow$
If $E \in Cn_\Sigma(Z)$ is such that E = FOR then, by **Theorem** *2, E is inconsistent. If E is inconsistent, by* **Theorem** *3, $T \cup Z$ is inconsistent.*
$\Leftarrow$
If $T \cup Z$ is inconsistent then, by **Theorem** *3, every $E \in Cn_\Sigma(Z)$ is inconsistent. If E is inconsistent then, by* **Theorem** *2, E = FOR.*

With these precedent formal elements, in this section we will characterize in **GMD** (1) Reiter's defaults, (2) Generalized Reiter's defaults, and (3), Reiter's systems, by means of modulations based on the notions of a base of provisos K and a relation between K and the absence condition B.

A Reiter's default can be captured in **GMD** as a non-standard rule $\langle A, \gamma, \{\beta_1, \ldots, \beta_n\}\rangle$ in a mds Σ, such that for every extension E of Σ, its base of provisos K is such that
(1) $K = E$,
(2) its condition C is: $\neg\beta_1, \ldots, \neg\beta_n \notin K$.

The more familiar presentation of a Reiter's default can be presented in our framework **GMD** as $\langle\{\alpha\}, \omega, \{\beta_1, \ldots, \beta_n\}\rangle$. Reiter writes a default as:

$$\frac{\alpha : \diamond\beta_1, \ldots, \diamond\beta_n}{\omega}$$

This structure could be read it as follows:
"If α, infer ω provided that our epistemic state is such that it is consistent to assume every β_i in the sequence $\beta_1, \ldots, \beta_n$".

Other kinds of default rules can be constructed with different modulations. For example, we can make a modulation by modifying condition C, as follows:

Differently from Reiter's default, a default based on non triviality can be captured in **GMD** as a non-standard rule $\langle A, \gamma, B\rangle$ in a mds Σ, such that for every extension E of Σ, its base of provisos K is such that
(1) $K = E$,
(2) its condition C is: $K \neq FOR$.

In order to combine non-monotonicity and paraconsistency, we will make now a different modulation in a default rule. This modulation is a combination between Reiter's defaults and the precedent non trivial defaults. The new modulation is a certain kind of combination of both conditions C of the two precedent kinds of defaults.

We consider that a Standard Non Trivial Default (SNT-Default) as an nsr $\langle A, \gamma, B\rangle$ in a mds $\Sigma = \langle \Lambda, T\rangle$, such that for every extension E of Σ, its base of provisos K is such that
(1) $K = E$,
(2) its condition C is: $\bigcup Cn_{\Lambda S}(B \cup K) \neq FOR$.
Where ΛS is the set of standard rules $[A, \gamma] \in \Lambda$.

Remark 5 Our above notion of default is called by us "Standard Non Trivial Default" (or "NTS-Default") because its modulation by means of 1 and 2, is based upon triviality and not upon contradiction and its condition C of non triviality is based upon the set of standard rules of Λ (it is different from Reiter's original proposal, here represented by a Reiter's default).

Even if we use SNT-Defaults, from the extensional point of view, the effect that is produced taking Λ as da Costa's classical calculus $C_0^{*=}$, is the same than that produced in a Reiter's default theory. What is most important for our goals, is that the change described in **Remark** 5 will permit, in a very natural way, the interaction between paraconsistency and default reasoning which will be developed in the next section.

To emphasize more clearly the way in which systems similar to those of Reiter can be varied in **GMD**, we will show each component of the systems as an item that can be modulated, also modifying the interaction of the different components within the system. For simplicity, in this presentation of **GMD** we will only introduce a class of modulations, namely, restrictions, but it is possible to implement modulations such as amplifications or couplings of various restrictions. We can now represent a default theory through an mds with certain particular modulations, as following:
A Reiter's system in **GMD** is a mds $\Sigma = \langle \Lambda, T \cup Z\rangle$ such that:

(A) $C_0^{*=} \subseteq \Lambda$ (i. e., in the system we encounter this particular logical restriction).
(B) Every $[A, \gamma] \in \Lambda$ is such that $\langle A, \gamma \rangle \in C_0^{*=}$ (i.e., a particular restriction of the structure of rule).
(C) Every nsr $\langle A, \gamma, B \rangle \in \Lambda$ is an SNT-Default (i.e., a particular restriction of structure of rule).
(D) Every extension $E \in Cn_\Sigma(Z)$ is such that $E = \bigcup_{i=0}^{\infty} E_i$, where for the family $\{E_i\}_{i \in \mathbf{N}}$, the following conditions hold: $E_0 = T \cup Z$.
For all E_i, where $i \geq 0$, $E_{i+1} = \bigcup Cn_\Lambda S(E_i) \cup \{\gamma \mid \gamma$ is the conclusion of a particular nsr $\langle A, \gamma, B \rangle \in \Lambda, A \subseteq \bigcup Cn_\Lambda S(E_i)$, and the relation C, between B and the respective base of provisos K of $\langle A, \gamma, B \rangle$, is fulfilled$\}$ (i.e., a particular restriction of the interaction logic-theory).
ΛS is the set of standard rules $[A, \gamma] \in \Lambda$.

We have proposed a very general structure of a rule, which supports modulations in two senses: specifying the base of provisos K, and specifying the condition C that relates B with K. And we also have proposed three additional kinds of modifications: on the supposed underlying logic respect to a theory, on the possible change of the theory T, and on the procedure of interaction between theories and its underlying logics (through construction of extensions of a theory by means of Cn_Σ). Together those modifications provide us with a formal framework to analyze the interaction between rules, theories and their underlying logics: we can vary some of the mentioned items with gradual modifications and, in this way, analyze the variation in relation to their effects in terms of different mds's Σ and their different sets of consequences.

Now we will present some results related with Reiter's systems.

Theorem 5 *A Reiter's system $\Sigma = \langle \Lambda, T \cup Z \rangle$ has an inconsistent extension E, iff $T \cup Z$ is inconsistent.*
Sketch of proof:
$\Rightarrow$
Let E^1 be inconsistent, then there exists $\alpha \in FOR$, such that $\{\alpha, \neg\alpha\} \subseteq E^1$. But, by condition D of a Reiter's system, every E is that $E = \bigcup_{i=0}^{\infty} E_i$. Finally, E^1 is the result of the function $Cn_{\Lambda S}$ when there is not another SNT-default to apply. Assume that $T \cup Z$ is consistent. Then, there exists an E_j^1 such that E_{j-1}^1 is consistent and E_j^1 is inconsistent. Consequently, there exists at least one subset of applied SNT-defaults $\langle A, \gamma, B \rangle \in \Lambda$ (SAD now on) such that $SAD \cup Cn_{\Lambda S}(E_{j-1}^1)$ is inconsistent and $SAD \cup Cn_{\Lambda S}(E_{j-1}^1) \subseteq E^1$. But this would imply that E^1 is FOR, because, as we said, E^1 is the result of the function $Cn_{\Lambda S}$. But, by construction, every applied SNT-default must comply with its condition C, namely, $\bigcup Cn_{\Lambda S}(B \cup E^1) \neq FOR$. Thus, our assumption is false. Therefore, $T \cup Z$ is inconsistent.

$\Leftarrow$
Since every extension $E \in Cn_\Sigma(Z)$ is such that $T \cup Z \subseteq E$, therefore, if $T \cup Z$ is inconsistent then E is inconsistent.

Theorem 6 *For every pair of extensions of a Reiter's system $\Sigma = \langle \Lambda, T \cup Z \rangle$, E^1 and E^2, $E^1 \cup E^2$ is inconsistent. In this case we say that E^1 and E^2 are orthogonal.*

Sketch of proof:
$T \cup Z$ is or a consistent or an inconsistent set.
In the first case, we know that, E^1 and E^2, are the result of the application of different SNT-defaults such that applying some of them imply the impossibility of applying others. This impossibility consists of non-compliance of the condition C. But this condition is only that the resulting E must not be the trivial set. In these situation, for every pair E^1 and E^2, $E^1 \cup E^2$ is trivial. But the trivial set is an inconsistent set. Therefore, in these conditions, for every E^1 and E^2, $E^1 \cup E^2 = FOR$.
Second case, we know that every extension E will be such that $Cn_{\Lambda S}(T \cup Z) \subseteq E$, but, in a Reiter system, ΛS is classical logic, then, $E = FOR$. Therefore, in this situation, for every E^1 and E^2, $E^1 \cup E^2$ is inconsistent.

Theorem 7 *If a Reiter's system $\Sigma = \langle \Lambda, T \cup Z \rangle$ has an inconsistent extension then it is its unique extension.*
Sketch of proof:
By **Theorem** *5, if has some inconsistent extension E^1, then $T \cup Z$ is inconsistent. And by construction, every extension E is such that $Cn_{\Lambda S}(T \cup Z) \subseteq E$. Then, in these conditions, $FOR \subseteq E$, for every E, i.e., in these conditions, every $E = FOR$. Therefore, in these conditions, every E^1 of Σ is such that for every other extension E of Σ, $E^1 = E = FOR$.*

Theorem 8 (Minimality of extensions) *If E^1 and E^2 are extensions of a Reiter's system, such that $E^1 \subseteq E^2$, then $E^1 = E^2$.*
Sketch of proof:
Suppose that $E^1 \subset E^2$. By construction, for every extension E, $E = \bigcup_{i=0}^{\infty} E_i$. Then if $E^1 \subset E^2$, there is a semiextension E_j^2, such that $E_j^2 = E^1$. Thus E^1 is not an extension at all. But this fact contradicts the initial condition that E^1 and E^2 are extensions. Therefore, $E^1 \not\subset E^2$. Given $E^1 \subseteq E^2$, then $E^1 = E^2$.

Theorem 9 (Global Non-monotonicity) *Not every non trivial Reiter's System $\Sigma = \langle \Lambda, T \rangle$ is such that for any $Z \subseteq FOR$, $Cn_{\Sigma}(\emptyset) \subseteq Cn_{\Sigma}(Z)$.*
Sketch of Proof:
Let $\Sigma = \langle \Lambda, T \rangle$, $\Lambda = C_0^{*=} \cup \{\langle \{\alpha\}, \gamma, \{\beta\} \rangle\}$, $T = \{\alpha\}$.
With these conditions, if we consider $Z = \{\neg\beta\}$ we would have the following results: $\bigcup Cn_{\Sigma}(\{\alpha, \gamma\}) \in Cn\Sigma(\emptyset)$ and, nevertheless, $\bigcup Cn_{\Sigma}(\{\alpha, \gamma\}) \notin Cn\Sigma(Z)$.

Theorem 10 (Relative Non-monotonicity) *Given a non trivial Reiter's System $\Sigma = \langle \Lambda, T \rangle$, not every extension E^{Σ} of Σ is such that there exists an extension $E^{\Sigma Z}$ of $\Sigma = \langle \Lambda, T \cup Z \rangle$, such that $E^{\Sigma} \subseteq E^{\Sigma Z}$.*
Sketch of Proof:
The interesting case:
Let $\Sigma = \langle \Lambda, T \rangle$, $\Lambda = C_0^{*=} \cup \{\langle \{\alpha\}, \gamma, \{\beta\} \rangle\}$, $T = \{\alpha\}$. Under these conditions, if we consider $Z = \{\neg\beta\}$ we would have the following results: $Cn_{\Sigma}(\emptyset) = \{\bigcup Cn_{\Sigma}(\{\alpha, \gamma\})\}$ and $Cn_{\Sigma}(Z) = \{\bigcup Cn_{\Sigma}(\{\alpha\})\}$

Let $\bigcup Cn_\Sigma(\{\alpha, \gamma\})$ *be* E^Σ *and* $\bigcup Cn_\Sigma(\{\alpha\})$ *be* $E^{\Sigma Z}$. *We can observe that in this case, it does not occur that* $E^\Sigma \subseteq E^{\Sigma Z}$.
The limit case:
Let $\Sigma = \langle \Lambda, T \rangle$, $\Lambda = C_0^{*=} \cup \{\langle\{\alpha\}, \gamma, \{\beta\}\rangle, \langle\{\alpha\}, \phi, \{\neg\gamma\}\rangle\}$, $T = \{\alpha\}$. *Under these conditions, if we consider* $Z = \emptyset$ *we would have the following results:*
$Cn_\Sigma(\emptyset) = \{\bigcup Cn_\Sigma(\{\alpha, \gamma\}), \bigcup Cn_\Sigma(\{\alpha, \phi\})\} = Cn_\Sigma(Z)$.
Let $\bigcup Cn_\Sigma(\{\alpha, \gamma\})$ *be* E^Σ *and* $\bigcup Cn_\Sigma(\{\alpha, \phi\})$ *be* $E^{\Sigma Z}$. *We can observe that in this case, it does not occur that* $E^\Sigma \subseteq E^{\Sigma Z}$ *either.*

Remark 6 **Theorems** 5, 6, 8, and **Theorem** 7, correspond to Corollaries 2.2, 2.3 and Theorems 2.4 and 3.3 in [17].

5 Building a Paraconsistent Default Theory on GMD

A default derivation is based on an observation of consistency by means of its correspondent base of provisos K and relation C. This makes it difficult to connect defaults and paraconsistent logics. Due to this, we presented a modified version of a non-standard rule: a NTS-Default. We need one more modification in order to allow the interaction between non-monotonicity and paraconsistency. We will use the definition of a Reiter's system with a change in its logical modulation. We can do that by means of a paraconsistent logic. Thus, we will assume, as an example of constituent of Λ, the axioms of C_1 of the paraconsistent calculi from da Costa's hierarchy $C_n^{*=}$, $1 \leq n \leq \omega$. As before, we will understand C_1 as a set of rules of the form $\langle A, \gamma \rangle$ (not to be confused with $[A, \gamma]$, which, as we said above, abreviates an ordered triple).

By modifying the base of provisos K and its corresponding relation between B and K, we have characterized a Reiter's system. We must do a little modulation in order to fulfill our aim. If we replace calculus C_0 with C_1 in da Costa's hierarchy, we will obtain a mds that can support the interaction between non-monotonicity and paraconsistency. Similarly to our presentation of Reiter's Systems, we will present the different components that have been modulated in the Reiter-da Costa systems.

Definition 6 (*Definition* 6) We define a Reiter-da Costa System (or an RC-system) as a mds $\Sigma = \langle \Lambda, T \cup Z \rangle$, where:
(A) $C_1 \subseteq \Lambda$ (i.e., a particular logical modification).
(B) Every $[A, \gamma] \in \Lambda$ is such that $\langle A, \gamma \rangle \in C_1$ (i.e., a particular restriction of structure of rule).
(C) Every nsr $\langle A, \gamma, B \rangle \in \Lambda$ is an SNT-Default (i.e., a particular restriction of structure of rule).
(D) Every extension $E \in Cn_\Sigma(Z)$ is such that $E = \bigcup_{i=0}^{\infty} E_i$, where for the family $\{E_i\}_{i \in \mathbf{N}}$, the following conditions hold: $E_0 = T \cup Z$.
For all E_i, where $i \geq 0$, $E_{i+1} = \bigcup Cn_\Lambda S(E_i) \cup \{\gamma \mid \gamma$ is the conclusion of a particular nsr $\langle A, \gamma, B \rangle \in \Lambda, A \subseteq \bigcup Cn_\Lambda S(E_i)$, and the relation C, between B and the respective base of provisos K of $\langle A, \gamma, B \rangle$, is fulfilled$\}$

Let's show some interesting results using Reiter-da Costa systems, instead of Reiter's systems.

Theorem 11 *Not necessarily, if E is an inconsistent extension of an RC-system Σ, $E = FOR$. Due to this, Σ is a paraconsistent mds.*
Sketch of Proof:
Let α be a badly-behaved member of FOR (α and $\neg\alpha$ can be both true). And let α, $\neg\alpha$, be the unique inconsistency in E. In these conditions we cannot obtain, by C_1, the result $E = FOR$.

Theorem 12 *If $\Sigma = \langle \Lambda, T \cup Z \rangle$ is an RC-system and $T \cup Z$ is inconsistent, then every extension E of Σ is inconsistent.*
Sketch of Proof:
By construction, every extension $E \in Cn_{\Sigma}(Z)$ is such that $T \cup Z \subseteq E$. Then, if $T \cup Z$ is an inconsistent set, so will be E.

Theorem 13 *For every pair of different extensions of an RC-system, E^1 and E^2, $E^1 \cup E^2$ is inconsistent. (Orthogonality)*
Sketch of Proof:
By construction, each extension is constructed by rules whose absence condition is based on non-triviality. Then, if E^1 and E^2 are different extensions, both belonging to $Cn_{\Sigma}(Z)$, then $E^1 \cup E^2$ generates the trivial set. This fact is possible only if they are mutually inconsistent extensions.

Theorem 14 *An RC-system $\Sigma = \langle \Lambda, T \cup Z \rangle$ has FOR as an extension iff $Cn_{C1}(T \cup Z) = \{FOR\}$.*
Sketch of Proof:
$\Rightarrow$
By construction, every extension E is constructed by the application of rules of C_1, and by the application of rules whose absence condition is based on non-triviality. Then, in these systems, the application of rules that are not in C_1 does not produce triviality. It follows that, if $E \in Cn_{\Sigma}(Z)$ and $E = FOR$ then the triviality was already in the consequences of the basic assumptions of the theory, i.e., $Cn_{C1}(T \cup Z) = \{FOR\}$.
$\Leftarrow$
If $Cn_{C1}(T \cup Z) = \{FOR\}$ then, by construction, $E_1 = FOR \cup \{\gamma \mid \gamma$ is the conclusion of a particular nsr $\langle A, \gamma, B \rangle \in \Lambda, \ldots\}$, then $FOR \subseteq E_1$; therefore, $FOR \subseteq E$. It follows than $E = FOR$.

Theorem 15 *For every pair of different extensions E^1 and E^2 of an RC-system $\langle \Lambda, T \cup Z \rangle$, $Cn_{C1}(E^1 \cup E^2) = \{FOR\}$, (i.e., $\bigcup Cn_{C1}(E^1 \cup E^2)$ is trivial). (Paraorthogonality)*
Sketch of Proof:
Similarly to **Theorem** *13, by construction every different extensions E^1 and E^2 are constructed by rules whose absence condition is based on non-triviality. Then, if E^1 and E^2 are different extensions, both belonging to $Cn_{\Sigma}(Z)$, then it follows that $Cn_{C1}(E^1 \cup E^2) = \{FOR\}$.*

Theorem 16 (Global Non-monotonicity)

(A) For some non trivial RC-System $\Sigma = \langle \Lambda, T \rangle$*, there exists a* $Z \subset FOR$ *such that* $Cn_\Sigma(\emptyset) \not\subset Cn_\Sigma(Z)$*.*
Sketch of Proof:
Let $\Sigma = \langle \Lambda, T \rangle$*,* $\Lambda = C_1 \cup \{\langle \{\alpha\}, \gamma, \{\beta\} \rangle\}$*,* $T = \{\alpha\}$*, and let* β *be a well-behaved formula (not badly behaved). Under these conditions, if we consider* $Z = \{\neg\beta\}$ *we would have the following results:*
$\bigcup Cn_\Sigma(\{\alpha, \gamma\}) \in Cn_\Sigma(\emptyset)$ *and, nevertheless,* $\bigcup Cn_\Sigma(\{\alpha, \gamma\}) \notin Cn_\Sigma(Z)$*.*

(B) For some non trivial RC-System $\Sigma = \langle \Lambda, T \rangle$*, there exists a* $Z \subset FOR$ *such that* $\bigcup Cn_\Sigma(\emptyset) \not\subset \bigcup Cn_\Sigma(Z)$*. The intuitive idea is that* $\bigcup Cn_\Sigma(Z)$ *represents the total set of available consequences (perhaps not simultaneously) from Z in* Σ*.*
Sketch of Proof:
Let $\Sigma = \langle \Lambda, T \rangle$*,* $\Lambda = C_1 \cup \{\langle \{\alpha\}, \gamma, \{\beta\} \rangle\}$*,* $T = \{\alpha\}$*, and let* β *a well-behaved formula. Under these conditions, if we consider* $Z = \{\neg\beta\}$ *we would have the following results:*
$\gamma \in \bigcup Cn_\Sigma(\emptyset)$ *and, nevertheless,* $\gamma \notin \bigcup Cn_\Sigma(Z)$*.*

Theorem 17 (Relative Non-monotonicity) *Given a non trivial RC-System* $\Sigma = \langle \Lambda, T \rangle$*, not every extension* E^Σ *of* Σ *is such that exists an extension* $E^{\Sigma Z}$ *of* $\Sigma = \langle \Lambda, T \cup Z \rangle$*, such that* $E^\Sigma \subseteq E^{\Sigma Z}$*.*
Sketch of Proof:

The interesting case:
Let $\Sigma = \langle \Lambda, T \rangle$*,* $\Lambda = C_1 \cup \{\langle \{\alpha\}, \gamma, \{\beta\} \rangle\}$*,* $T = \{\alpha\}$*, and let* β *a well-behaved formula. Under these conditions, if we consider* $Z = \{\neg\beta\}$ *we would have the following results:* $Cn_\Sigma(\emptyset) = \{\bigcup Cn_\Sigma(\{\alpha, \gamma\})\}$ *and* $Cn_\Sigma(Z) = \{\bigcup Cn_\Sigma(\{\alpha\})\}$
Let $\bigcup Cn_\Sigma(\{\alpha, \gamma\})$ *be* E^Σ *and* $\bigcup Cn_\Sigma(\{\alpha\})$ *be* $E^{\Sigma Z}$*. We can observe that in this case it does not occur that* $E^\Sigma \subseteq E^{\Sigma Z}$*.*

The limit case:
Let $\Sigma = \langle \Lambda, T \rangle$ *,* $\Lambda = C_0^{*=} \cup \{\langle \{\alpha\}, \gamma, \{\beta\} \rangle, \langle \{\alpha\}, \phi, \{\neg\gamma\} \rangle\}$*,* $T = \{\alpha\}$*. Under these conditions, if we consider* $Z = \emptyset$ *we would have the following results:*
$Cn_\Sigma(\emptyset) = \{\bigcup Cn_\Sigma(\{\alpha, \gamma\}), \bigcup Cn_\Sigma(\{\alpha, \phi\})\} = Cn_\Sigma(Z)$*.*
Let $\bigcup Cn_\Sigma(\{\alpha, \gamma\})$ *be* E^Σ *and* $\bigcup Cn_\Sigma(\{\alpha, \phi\})$ *be* $E^{\Sigma Z}$*. We can observe that in this case it does not occur that* $E^\Sigma \subseteq E^{\Sigma Z}$ *either.*

Remark 7 **Theorems** 11, 14, 15 and 16 are different results from that of [17]. **Theorem** 14 is analogous to Reiter's theorem about an inconsistent initial W. Mainly, in this new kind of doxastic system (RC-System), the differences in results are because of the fact that inconsistency does not imply triviality.

6 Explanations in the GMD Formal Framework

In the present section, we will use our abstract notion of rule and the constructed environment **GMD** that deals with theories and its underlying logics, to capture a minimal concept of explanation. This concept of explanation, we think, will deal with some of the main objections discussed in the main related debates in Philosophy of Science. In order to do that, we will use the Reiter-da Costa systems and begin constructing some new functions. We will begin with the analysis of three conditions in explanations in order to motivate our use of new modulations.

The notion of SNT-Defaults reflects the idea of explanation as an argumentative structure. Two difficulties remark in the debate on explanation in Philosophy of Science, which we can model with this structure, are the possibility of low probability arguments as explanations and the possibility of arguments with unspecified exceptions as arguments. Both difficulties have been seen in the philosophical literature of explanation as definitive obstacles to the construction of an argumentative formal model of explanation. If default structures capture a minimal idea of argument, and this in turn could capture a minimal idea of explanation then we may be able to represent minimal explanations based on default structures. Besides, not only there are counterexamples to the demand of a deductive relation in argument-base explanations (as we so in our discussion of AC4), there are also similar counterexamples against the need for a probabilistic relation in them. There is a great diversity of inferential relations that are non deductive and not even probable. For instance, we may mention plausibility and tipicity relations.[13] Due to this it would be interesting if the inferential relation assumed in an explanation remains open to variation that allowed to represent such different inferential relations.

On the other hand, the demand for truth in the premises of an explanation is also very restricted (as we have seen in the discussion about condition AC1 at the beginning of this paper). This demand would exclude historically important cases in Science, for instance, that of Phlogiston in the explanation of combustion. We can trust scientific theories to be generally in state of flux. This suggest a representation of scientific explanation that skirts the truth restriction on the premisses. Even more, we should avoid expecting all explanations to start from a consistent premise set. And this is so because explanations can not be evaluated in isolation with respect to the doxastic systems with which they interact and, oftentimes, it is difficult to asses whether those doxastic systems yield no contradictions. One way to represent this later possibility is to held doxastic systems as Reiter-da Costa systems. One may represent the possibility of falsehood in the *explanans* addressing the connection between *explanans* and *explanandum*, so to say, indirectly, and this maybe achieve maintaining this connection as a derivation relation that does not assumes a true left side. Thus, the pluses of the *explanans* depend only under general inferential or justificatory capabilities, not including the demand for truth. We shall explore this possibility of representation in this section.

In our model AC1–AC4 are desirable but non-essential characteristics of an explanation. We want to allow something to be called an explanation even if it violates one or more of this conditions. In the way of an elucidation of the concept of explanation, according to the reflections suggested by PS1–PS4 and in this section, our basic idea in this work is that an argument is an explanation if the facts supposed in its *explanans* allow to establish an explicative relevance relation between some

[13]See [13].

of these supposed facts and the fact supposed in the *explanandum*. We consider both *explanans* and *explanandum* as propositions. These propositions are assuming facts. An explanation is a relationship between a set of propositions (*explanans*) and another proposition (the *explanandum*). Differently, we call "explanatory relevance relation" to a relationship that happens between facts: between the facts supposed in the *explanans* and the fact supposed in the *explanandum*. The precise nature of the explanatory relevance relation is an unsolved problem in Philosophy of Science, but we will assume for the time being a minimal notion of explanation (which includes a distinction between explanation and explanatory relevance relation) that we think contributes to solve this kind of problems in the modeling of scientific explanations.

Of course, to say that an explanatory relation is a relation between facts is not very precise. Our proposal in this paper does not commit us to one of these explanatory relationships in particular. We just try to clarify how the distinction between an explanation and the explanatory relevance relation it supposes helps to understand the interaction between explanations and their theoretical contexts, using a non-monotonic and paraconsistent formal environment.

We do not need to reject (as non necessary) every feature of explanation. Argument-base explanations tend to demand stronger than normal justifications conditions. For instance, a coherence is demanded in the inferential mechanism itself specially in the interaction of the rules used to build a belief set. We will analyze this in more detail in the case of justifications involving non standard rules. For instance, the activation of defaults require only the consistency of certain truths with respect to the final extension. But, in some cases, we might be interested, for a variety of reasons, in the consistency between the propositions themselves deemed comparable in the activated rules. We might want all justifications comparable with the final extensions to be comparable among themselves. We shall explore this special requirement below.

Our abstract notion or rule allows for particular instances of modulation of the mechanism of inference inside specific doxastic systems.

Let us suppose the next three situations under involving absence conditions.
(AbCondS0) Suppose Example EX, and suppose that Joseph is reflecting about the situation. He considers completely true four things besides P1 and P2:

(P3) John used to put the arsenic in a jar similar to the jar of sugar,
(P4) John used to have dangerous enemies,
(P5) If P3 then, (C1) John was a very irresponsible person,
(P6) If P4 then, (C2) John had a complicated life.

With this set of beliefs at his disposal, Joseph infers with a lot of certainty: (C1) and (C2).

(AbCondS1) Suppose Example EX, and suppose that Joseph is reflecting about the situation. He considers completely two things besides P1 and P2: P3 and P4 (but not necessarily P5 and P6). Furthermore Joseph engages in two following inferential processes:

(I1) Based on P3, and assuming that:
(As1) John made peace with his dangerous enemies

Joseph infers by some non-standard rule that:
(C3) John died accidentally.
(I2) Based on P4, and assuming ¬(As1), Joseph infers by some non-standard rule ¬(C3).

(AbCondS2) Suppose Example EX, and suppose that Joseph is reflecting about the situation. He considers completely true two things besides P1 and P2: P3 and P4 (but not necessarily P5 and P6). Furthermore Joseph engages in two following inferential processes:
(I3) Based on P3, and assuming (As1), Joseph infers by some non-standard rule (C3).[14]
(I4) Based on P4, and assuming ¬(As1), Joseph infers by some non-standard rule (C2) John had a complicated life.

AbCondS0 describes a classical situation. Joseph infers from true premises by simple Modus Ponens independently of any assumption. He is using standard rules (with $B = \emptyset$). He does not need to check absence conditions.

The situation AbCondS1 is very different. In this situation Joseph is inferring by non-standard rules (with $B \neq \emptyset$), so he needs to check absence conditions. Note also that Joseph has alternative inferential processes in this situation. He rationally cannot maintain the inference I1 jointly with I2. We have, then, two different extensions of his base belief set. If Joseph has a preference for one of these extensions, he needs rationally to cancel the other one because of the contradictory conclusions (C3 and ¬C3) of I1 and I2. In certain cases, when we apply a non-standard rule (nsr), we want to check the consistency between the set B of nsr and the set of consequences.[15]

In the situation AbCondS2 we also have non-standard rules. And we have another interesting situation. Let us observe that (C3) and (C2) are not contradictory conclusions. We could rationally maintain both sentences at the same time. Nevertheless, the provisos that lead us to infer (C3) and (C2) contain contradictory sentences. In order to infer (C3) Joseph assumes (As1), but to infer (C2), he assumes ¬(As2). Technically, he can accept (C2) and (C3) without contradiction. Nevertheless, Is it advisable to assume that both provisos hold? This is an interesting question whose answer might depend on our specific theoretical interest for using models that include provisos. In what follows we will explore the approach that would rather dispense with a incoherence sets of activated provisos.

An answer to this question depends upon our conception of assumption in the context of a proviso. The proviso used to infer (C2) is not compatible with that used to infer (C3). If there were no other justifications and we want to have a consistently justified set of beliefs, we should not maintain (C2) and (C3).

Departing from [17], the set of sentences that a doxastic system uses to check the provisos in B can be different depending on the kind of rationality in question. It is convenient to maintain a general formulation of the absence conditions (B) in a non-standard rule (a rule with $B \neq \emptyset$), such that it allows for different ways of

[14] As the reader might notes, I1 and I3 are the same. The difference between the situation AbCondS1 and AbCondS2 is with respect to I2 and I4.

[15] We leave for another paper the exploration of the epistemic significance of such cases.

interaction between the beliefs that we consider with a high level of certainty and the beliefs we consider with a low level of certainty.

In AbCondS1 it is sufficient to check the absence condition (the provisos) of each inference I1 and I2 in the set of final consequences, the final extension. In this way, if we find a contradiction, we cancel the inference of one of I1 or I2. In AbCondS2, this strategy will not be sufficient. We will need to check the absence condition of its inference I3 and I4 in the set of final consequences but also to check that those absences conditions do not conflict with any prerequisite or provisos in any of the rules used to generate the set of final consequences. Even more we may need to check each proviso used against the set of prerequisites and provisos used as a whole. Basically, in this situation, when we apply a non-standard rule, as we are supposing here, we need to check consistency inside the set of consequences Γ in union with the absence conditions of that non-standard rule, in union with the absence conditions of any other rule used in order to construct Γ.

Our analysis of AbCondS0, AbCondS1, and AbCondS2 in relation with the absence condition of a rule, sheds light on an important modularity of mds. We may vary the absence condition of a rule and, by this change, allow for variation in the way of constructing the set of consequences. By means of these possible variations, we can model different kinds of consequences. Another interesting case of this modulation is when the same set of consequences results from the variations but each time justified differently.

If we want to capture certain kinds of reasoning coherence, we will need to capture the interaction between provisos of different rules used in a doxastic system. This can lead us to consider, as a reference in the application of a particular rule, not only the extension that is being constructed, but also the way in which that extension is being put together. Before we analyze other important conditions to incorporate in a model of explanation, let's offer an adequate representation of these cases of coherence of provisos by means of a certain new modulation.

Given an RC-System $\Sigma = \langle \Lambda, T \cup Z \rangle$, we have our previous function Cn_Σ that returns the set of all extensions E of Σ. Now, we will assume $D_{\subseteq \Lambda} = \{d_1, d_2, \ldots, d_n\}$ as the set of SNT-Defaults in Σ. We then use a function BP that takes an extension $E \in Cn_\Sigma$, and returns the set of sequences $\{s_1, s_2, \ldots, s_m\}$ such that, for all s_i:

(a) $s_i = \langle d_{i_1}, d_{i_2}, \ldots, d_{i_k} \rangle$ is a sequence in which some elements (not necessarily all the elements) of $D_{\subseteq \Lambda}$ can be applied to produce E.
(b)
$E_0 = T \cup Z$,
$E_1 = \bigcup Cn_\Lambda(E_0) \cup \{\omega_1 \mid \omega_1$ is the conclusion of d_{i_1} in $s_i\}$,
$E_2 = \bigcup Cn_\Lambda(E_1) \cup \{\omega_2 \mid \omega_2$ is the conclusion of d_{i_2} in $s_i\}$,
⋮
$E_k = \bigcup Cn_\Lambda(E_{k-1}) \cup \{\omega_k \mid \omega_k$ is the conclusion of d_{i_k} in $s_i\}$,
$\bigcup Cn_\Lambda(E_k) = E$.

Remark 8 We could construct one and the same extension E with different sequences of defaults. The number of defaults in $D_{\subseteq \Lambda}$ is not necessarily the same number of sequences to construct E. Thus, not necessarily $n = m$. It is also important to note that these different sequences can be composed of elements of different subsets of $D_{\subseteq \Lambda}$. Therefore, it is not necessary that $n = k$. A limit case to construct an extension could be the case in which $k = 0$.

We can think of $BP(E)$ on Σ as the set of building paths using Σ towards E.

If we consider B_n the absence condition of d_n, and in general, B_i the absence condition of d_i in a building path s_k, we can use now a function J such that applied to a building path s_k of Σ for the construction of an extension E, i.e., $J(s_k)$, gives us the set of every $\beta \in B_i$ of every d_i in s_k.[16] Basically, the function $J(s_k)$, where s_k is a building path to construct E, produces the set of all provisos used to construct E by means of s_k.

With these concepts at hand we are able, in certain particular cases, to distinguish two extensions such that $E^p = E^q$, because of two different building paths $s_p \neq s_q$. Then, given a particular doxastic system Σ, for all $E \in Cn_\Sigma$ and each $s_k \in BP(E)$, we may define a tuple $\langle E, s_k \rangle$.

Including these new notions in **GMD**, we can construct a new modulation that produces a new default structure. We define a Standard Non Triviality Default with Proviso coherence (SNTP-Default) as an nsr $\langle A, \gamma, B \rangle$ in a mds $\Sigma = \langle \Lambda, T \cup Z \rangle$, such that

(1) The base of provisos K of $\langle E, s_E \rangle$ is $K = E \cup J(s_E)$,

(2) The condition C is: $Cn_{\Lambda S}(B \cup K) \neq \{FOR\}$.

Where ΛS is the set of all standard rules $[A, \gamma] \in \Lambda$.

We can now define another kind of doxastic system in **GMD**. In order to do that, we take the definition of an RC-System, and we will only modulate the non-standard rules contained in its corresponding Λ.

Definition 7 (*Definition* 7) We define a Reiter-da Costa System with provisos coherence (or an RCP-system) as a mds $\langle \Lambda, T \cup Z \rangle$, where:

(A) $C_1 \subseteq \Lambda$.

(B) Every $[A, \gamma] \in \Lambda$ is such that $\langle A, \gamma \rangle \in C_1$.

(C) Every nsr $\langle A, \gamma, B \rangle \in \Lambda$ is an SNTP-Default (i.e., a particular restriction of structure of rule).

(D) Every extension $E \in Cn_\Sigma(Z)$ is such that $E = \bigcup_{i=0}^{\infty} E_i$, where for the family $\{E_i\}_{i \in \mathbf{N}}$, the following conditions hold: $E_0 = T \cup Z$.

For all E_i, where $i \geq 0$, $E_{i+1} = \bigcup Cn_\Lambda S(E_i) \cup \{\gamma \mid \gamma$ is the conclusion of a particular nsr $\langle A, \gamma, B \rangle \in \Lambda, A \subseteq \bigcup Cn_\Lambda S(E_i)$, and the relation C, between B and the respective base of provisos K of $\langle A, \gamma, B \rangle$, is fulfilled$\}$.

With these changes in a doxastic system, we can then observe that different derivations of the same $\phi \in FOR$ may exist in the same extension E^q, depending upon different pairs $\langle E^q, s_{q_1} \rangle$, $\langle E^q, s_{q_2} \rangle$, in a same RCP-system Σ. We will write in symbols that a derivation of ϕ from Z in an RCP-system Σ exists by means of $\langle E, s \rangle$ as:

---

[16]Although ours is different, the J function is inspired on a similar function J (applied on one particular default) that appears in Sect. 3.3 of [12].

$Z \vdash_{\Sigma_{Es}} \phi$.
We can consider ϕ as a disjunctive syntactic consequence from Z in an RCP-system $\Sigma = \langle \Lambda, T \rangle$ iff $\phi \in \bigcup Cn_{\Sigma}(Z)$.
And we can consider ϕ as a conjunctive syntactic consequence from Z in an RCP-system $\Sigma = \langle \Lambda, T \rangle$ iff $\phi \in \bigcap Cn_{\Sigma}(Z)$.

These modifications, SNTP-Defaults and RCP-Systems, are sufficient to capture the interaction between paraconsistency and non-monotonicity in an argumentative schema relative to a particular doxastic system. With these kinds of rules, we can capture in **GMD** a basic notion of explanation, a minimal notion of argumentative explanation that is able to include derivability with low probability as explanation. In addition, this notion checks for coherence between the assumptions we usually make in the construction of an explanation. This last characteristic entails a distinction between extensions and non deductive inferences involved in their construction. We understand extensions of a doxastic system set theoretically. And we are understanding building paths used to construct extensions (the non-deductive inferences used to make extensions), as sequences of applied rules to construct an extension from a particular doxastic system.

Since the aim of **GMD** is to analyze the interaction between explanations and, as we said before, a minimal version of different contexts, we need to link the minimal conditions of explanations with variations in the components of different possible doxastic systems. We will do this through an meta-schema that does not assume any particular doxastic system. We will use the general structure of a rule also in the construction of this meta-schema. In addition, we ought to note, that formal tools in **GMD** allow us only to deal with situations related to problems about inference and problems about context (problems of kind I and III). Problems of kind II, related to the notion of explicative relevance, are not yet included in our representation. Taking all this into consideration, the meta-schema we will construct in **GMD** will need to have components to capture:
(MS1) Different derivability relations,
(MS2) Different doxastic systems,
(MS3) A particular notion of possibility that allows us to represent the paraconsistent environment not inside of a doxastic system, but in the interaction between explanations and doxastic systems,
(MS4) Different explicative relevance relations.

Our unit of analysis is not the explanation in an absolute sense, but a candidate for explanation in relation to a particular doxastic system. This allows us to analyze the interaction between explanations and doxastic systems. An important feature of our proposal is that it allows to analyze if a set of statements can be considered an *explanans* for a phenomenon with respect to a doxastic system even though that set of statements does not belong to the doxastic system in question. This implies being able to implement an analysis on the possibility that a set of statements is integrated into a particular doxastic system producing an explanation, but at the same time maintaining this integration as a mere possibility. Given that the kind of systems that we use for the notion of explanation is Reiter-da Costa system, this possibility of integration must also be treated in a paraconsistent way.

In this line of thought, we will exemplify MS1 and MS2 with the help of NTP-Systems. In order to capture MS3, we define now a notion of paraconsistent possibility for **GMD**.

Definition 8 (*Definition* 8) We will consider that a formula $\phi \in FOR$ is paraconsistenly possible in a doxastic system $\Sigma = \langle \Lambda, T \rangle$, in symbols, $\Diamond_{\Sigma}\phi$, iff exists an extension E of Σ with a building path s_k, such that:
$Cn_{C1}(E \cup \{\phi\} \cup (s_k)) \neq \{FOR\}$.

Another and more illustrative way to see this last definition of paraconsistent possibility is the following. Respect to some doxastic system $\Sigma = \langle \Lambda, T \rangle$, we have that $\Diamond_{\Sigma}\alpha$ iff exists an extension E with an s_k of Σ, such that:

$\alpha \in \bigcup Cn_{C1}(E \cup J(s_k))$ OR $\neg\alpha \notin \bigcup Cn_{C1}(E \cup J(s_k))$

If we do not suppose the truth of the last disjunction, we will have four cases:

(i) Redundant Case. When α is in $\bigcup Cn_{C1}(E \cup J(s_k))$ and is a well-behaved formula: $\alpha \in \bigcup Cn_{C1}(E \cup J(s_k))$ AND $\neg\alpha \notin \bigcup Cn_{C1}(E \cup J(s_k))$.

(ii) Trivial Case. When α is not in $\bigcup Cn_{C1}(E \cup J(s_k))$ and is well-behaved formula: $\alpha \notin \bigcup Cn_{C1}(E \cup J(s_k))$ AND $\neg\alpha \in \bigcup Cn_{C1}(E \cup J(s_k))$

(iii) Paraconsistent Case. When $\neg\alpha$ is in $\bigcup Cn_{C1}(E \cup J(s_k))$ and is a bad-behaved formula: $\alpha \in \bigcup Cn_{C1}(E \cup J(s_k))$ AND $\neg\alpha \in \bigcup Cn_{C1}(E \cup J(s_k))$

(iv) Ignorant Case. When neither α nor $\neg\alpha$ are in $\bigcup Cn_{C1}(E \cup J(s_k))$: $\alpha \notin \bigcup Cn_{C1}(E \cup J(s_k))$ AND $\neg\alpha \notin \bigcup Cn_{C1}(E \cup J(s_k))$.

α is paraconsistently possible in the cases i, iii and iv. The trivial case makes the disjunction false. We will write $\Diamond_{\Sigma}(A)$ to denote that every $\phi \in A$ is paraconsistently possible in Σ.

At the beginning of our paper we presented a couple of demands that are usually made in the discussion about what should be included in a model of scientific explanation: AC2, that there must be a causal relation, AC3, that the *explanans* must describe the causal relation that actually produced the *explanandum*. On that occasion we mentioned that both conditions should be analyzed in contrast to the fact that many senses of causal relation could be found that could satisfy some intuitions of what should be required as explanatory relevance relation. It would be convenient if a minimal model of explanation had the possibility of modulations of the relevance relation required according to different ways of understanding what would be a good explanation. However, the required explanatory relevance relation should not be confused with the relation of derivation present in an explanation. The first is a relation between phenomena, the second, between statements. We will try a characterization of the explanatory relevance relation that includes these considerations.

Now, with the aim of capturing MS4, we will introduce a notation to represent a explanatory relevance relation between phenomena. The intuitive idea of a relevant explanatory relation from the explaining phenomenon d to the phenomenon to be explained e is a relation in which d gives account of the existence of e. d may have this kind of relation with e through a chain in which every link is a phenomenon with a relevant explanatory relation to other links.

We say that:
(a) $[chain_n]$ is a chain of n linked phenomena, $chain_n$ is the set of elements of $[chain_n]$, and $D(chain_n)$ is the propositional description of the elements of $chain_n$.
(b) If $\{e_1, e_2, e_3, \ldots, e_{n-1}, e_n\}$ are the set of linked phenomena that make up $[chain_n]$, then the set of edges of $[chain_n]$ is $\{(e_1, e_2,), (e_2, e_3,), \ldots, (e_{n-1}, e_n)\}$.
(c) If $[chain_n]$ is an initial subchain of $[chain_{n+1}]$, then $e_w \bowtie [chain_n]$ means that e_w is the last linked phenomenon in $[chain_{n+1}]$.
(d) $D(e), D(chain_n), D([chain_n]), D(e_w \bowtie [chain_n])$ stand respectively for a description of some phenomenon e, a description of the elements of $[chain_n]$, a description of $[chain_n]$ itself, and a description of e_w being the last element in $[chain_{n+1}]$.

We are ready now to model in **GMD** how a minimal notion of explanation works as an argument, taking into consideration the problems mentioned in I, II, III.

Definition 9 (*Definition* 9). We say that there is a relation of explanation between a set of propositions A and the phenomena e relative to an RCP-System $\Sigma = \langle \Lambda, T \rangle$ iff there is an ordered pair $\langle E^{\Sigma}, s_k \rangle$ in Σ and a pair $\langle E^{\langle \Lambda, T \cup A \rangle}, s_j \rangle$ in $\langle \Lambda, T \cup A \rangle$ such that:

(Ex1) $A \vdash_{E^{\Sigma}_{s_k}} D(chain_n)$.

(A description of the elements of the set of phenomena $chain_n$, $D(chain_n)$, is derivable from A, in Σ through s_k).

(Ex2) $\Diamond_{E^{\langle \Lambda, T \cup A \rangle}_{s_j}} D(e \bowtie [chain_n])$.

(It is paraconsistently possible with respect to $\langle E, s_k \rangle$ of $\langle \Lambda, T \cup A \rangle$ that e be the next member of the chain $[chain_n]$).

(Ex3) $\Diamond_{E^{\Sigma}_{s_k}} A$.

(Every element of A is paraconsistently possible respect to $\langle E, s_k \rangle$ of Σ).

(Ex4) $\Diamond_{E^{\langle \Lambda, T \cup A \rangle}_{s_j}} D(e)$.

(The description of e is paraconsistently possible in E_{s_k} of $\langle \Lambda, T \cup A \rangle$.

Note that in this approach the explanation relation holds between a set of propositions A and a phenomenon e through an explanatory relevance relation between the phenomenon described by the *explanans* A (in the description of the chain) in the theoretical context of the RCP-System Σ. A pair $\langle A, e \rangle$ can be evaluated in **GMD** in order to determine if there is a relation of explanation between its elements, under some RCP-System. Let us note too that the connection between the *explanans A* and the *explanandum e* is given by way of a description of the elements of some phenomena chain. Finally, the notion of possibility is introduced in order to keep some coherence with the postulates governing the RCP-Systems, and the evaluation outside these systems.

This way, in **GMD** we can distinguish three important elements in the model: the explanatory relevance relation, the relation of explanation and the explanation. Explanatory relevance relations are particular relations supposed to hold between phenomena. Some doxastic systems suppose them as a condition for something to be an explanation. A relation of explanation is a relation supposed to be satisfied between a set of propositions and a particular phenomenon. Relations of explanation are understood as constituted by inferential relations involving doxastic systems, but also are constituted by the fulfillment of the relevance relation. Finally, explanations are argument schemas that satisfy the extra-epistemic conditions that depend upon the fulfillment of a relation of explanation in the context of a particular doxastic system. Fulfillment of the relation of explanation depends, itself, among other things, upon the fulfillment of an explanatory relevance relation, in the context of a particular doxastic system too. This way, we can identify, from **GMD**, pairs of sets of propositions, descriptions of a phenomenon, $\langle A, D(e)\rangle$, as candidates to be explanations. These candidates can be expressed in **GMD** independently from doxastic systems. Interactions between these pairs of candidates can be compared with different doxastic systems, in order to determine whether the candidates can be considered explanations with respect to a particular doxastic system. This will depend on whether a relation of explanation between a different pair (the pair constituted by the set of propositions A and the phenomenon referred to by the description above, i.e. $\langle A, e\rangle$) is satisfied. The satisfaction relation between the members of the pair described involves an inferential relation and an explanatory relevance relation, connected one with the other in the same doxastic system.

An explanation can be represented too by a meta-argument (or a meta-schema), external to the epistemic systems, constituted by the elements Ex1–Ex4 just mentioned in the precedent lines.

A pair $\langle A, D(e)\rangle$ is an explanation with respect to $\Sigma = \langle \Lambda, T\rangle$ in **GMD**, iff there is a pair $\langle E, s_k\rangle$ in Σ such that between A and e the relation described by the following schema holds:

(EX) For every $\Sigma = \langle \Lambda, T\rangle$, phenomena e, set of formulae A, and chains $n, n+1$:

$$\frac{A \vdash_{E^{\Sigma}_{s_k}} D(chain_n) : \Diamond_{E^{\Sigma}_{s_k}} A \Diamond_{E^{\langle \Lambda, T\cup A\rangle}_{s_j}} D(e) \Diamond_{E^{\langle \Lambda, T\cup A\rangle}_{s_j}} D(e \bowtie [chain_n])}{D(e)}$$

which can be read as follows:
"If a description of the chain n can be derived from the set A in some context E_{s_k} of $\Sigma = \langle \Lambda, T\rangle$, A is paraconsistently possible in the context E_{s_k} of Σ, and the next two items are paraconsistently possible, in some context E_{s_j} of $\langle \Lambda, T \cup A\rangle$,
(i) the description of e, (i. e., $D(e)$),
(ii) the description of a chain $n+1$, in which e is the last edge (assuming the chain n as the initial subchain of chain $n+1$),
then, you may infer the description of e (i. e., $D(e)$) in the context E_{s_j} of $\langle \Lambda, T \cup A\rangle$."

If, under these conditions, $D(e)$ may be inferred, we can say that A explains e in the context of Σ.

This way, we could suppose that a criteria in **GMD** to identify pairs $\langle A, D(e)\rangle$ as explanations respect to a given doxastic system is that these pairs satisfy the schema EX. **GMD** can check whether a given pair $\langle A, D(e)\rangle$ would function as an explanation with respect to a particular doxastic system Σ, even if there is no extension E in Σ such that $A \cup \{D(e)\} \subseteq E$.

We can observe some interesting properties relative to this formal framework **GMD** to represent interactions between explanations and doxastic systems. Let A, A' be different sets of propositions; e a phenomenon; $[chain_m]$, $[chain_n]$ different chains of phenomena; and Σ, Σ' different RCP-Systems.

(e1) If there is a relation of explanation between A and e relative to Σ, and A is true (that is, every element in A is true), then this supports the truth of $D(e)$ in the context of Σ.
(e2) It is possible for A and $D(e)$ to be true in Σ, without a relation of explanation holding between A and e.
(e3) The fact that A keeps a relation of explanation with e in the context of Σ does not depend upon the fact whether A is true or not.
(e4) The fact that A has a relation of explanation with e in the context of Σ by means of $[chain_m]$ does not imply that extra-systematically (that is to say, outside Σ and A) the explanatory relevance relation between e and the phenomena that make true the elements in A, actually holds, that is, it does not imply the actual existence of a relation between e and $[chain_m]$.
(e5) Sometimes $\langle A, D(e)\rangle$ could be a *explanans*-*explanandum* pair (that is, it could be a relation of explanation) in the context of Σ, while it might not have this relation in the context of Σ'.
(e6) Sometimes $D(e)$ could be explained by A, and by A' too, in the context of the same Σ.
(e7) Sometimes $D(e)$ could be explained by $[chain_m]$ and by $[chain_n]$, in the context of the same Σ.
(e8) Sometimes, if $\langle A, D(e)\rangle$ is a *explanans*-*explanandum* pair, it could use $[chain_m]$ as a part of the explanation but it could also use $[chain_n]$, in the context of the same Σ.
(e9) Sometimes if $\langle A, D(e)\rangle$ is a *explanans*-*explanandum* pair in the context of Σ, and the information of Σ grows, then, even when the prerequisite of EX might be satisfied for Σ, it could happen anyway that, for this pair, the provisos no longer hold.
(e10) Sometimes, if $\langle A, D(e)\rangle$ is a *explanans*-*explanandum* pair in a given context Σ, when information grows the prerequisite does not keep its nexus with the explanatory relevance relation anymore (we cannot conclude $D(chain_n)$ from A). Then, $\langle A, D(e)\rangle$ no longer is a *explanans*-*explanandum* pair.
(e11) Sometimes, if $\langle A, D(e)\rangle$ is a *explanans*-*explanandum* pair in a given context Σ, when information grows, even if the prerequisite keeps its nexus with the explanatory relevance relation, the chain turns out not to be paraconsistently possible after the information growth and then, $\langle A, D(e)\rangle$ cannot be a *explanans*-*explanandum* pair either.
(e12) Sometimes, in a given context Σ, if $\langle A, D(e)\rangle$ is a *explanans*-*explanandum* pair, when information grows we cannot infer the description $D(e)$ from A anymore.

This formal frame **GMD** to model explanation puts together formal tools from paraconsistent and non-monotonic logic. Also, e1–e12 suggest that **GMD** may help to solve some important objections in Philosophy of Science and to build an argumentative model of scientific explanation. The model presents a version of the interactions between explanations and their theoretical contexts by means of representing minimal relations between components of the explanations.

7 Concluding Remarks

GMD is a formal framework to analyze the interaction between rules and a minimal conception of context (doxastic systems). The context is composed by a set of beliefs (a minimal idea of a theory) in interaction with an inferential engine (a logic). Our abstract characterization of rules allows us to have a wide variety of rules and kinds of interaction with several parts of a context, by means of our concept of modulation. We think that the formal environment inside **GMD** could contribute to clarify the combination and interaction between various logics and between their particular components.

Identifying explanations with argumentative schemas, **GMD** is able to represent in an abstract way explanations that take into account both paraconsistent and default forms of argumentation. The distinction between explanations, relations of explanation and explanatory relevance relations, allows an open doxastic view of the notion of explanation, such that we can understand explanations as theoretical proposals that not necessarily correspond to reality, but have specific properties that clearly establish connections with arguments and with belief systems. With this distinction at hand, we are also able to represent the interaction of particular cases of explanations with a context (which we represent, in a minimalistic way, as a simple doxastic system). This interaction has paraconsistent and non monotonic properties. The resulting model constitutes a theoretical proposal about a minimal notion of explanation, emphasizing the elucidation of the interaction between its internal components and of the interaction between the explanation and its context (what we could call its "contextual *explanans*"). We think that this proposal helps to solve some important representation problems in the theory of explanation from the point of view of Philosophy of Science. It will be interesting to explore the possibility of modeling with **GMD** different types of explanations in the literature, and to make some comparisons of their advantages or disadvantages. In particular, with respect to types of explanation of events.[17]

GMD provides the possibility to modify both internal items inside a doxastic system and also internal conditions of rules (modulation). A possibly fruitful future research would be to develop a broad concept of modulation and of its relation to the notions of translation[18] and combination of logics. In addition, modulation seems to be a promising tool for a unified study of many aspects of scientific explanation, and it seems applicable to other areas as well, such as Argumentation Theory, Discourse Analysis, etc.

The meta-schema proposed to model argumentative explanations also could be developed with different objectives than those of modeling explanations. We hope, for example, that the meta-schema will prove helpful in the development of criteria for the introduction of defaults in doxastic systems.

Many open problems remain. For instance, we characterize relevance in explanation as a chain structure that would be able to capture a part of the usual notion

---

[17]For example, in textbooks such as [14], or in several of the approaches presented in [18].

[18]In particular, translations in paraconsistent logic. For example, [5].

of causal relation, and also other possible notions of explanatory relevance relations (such as inferential relations, supervenience, etc.). We decided to leave open the possibility of other relations and that is a reason for this generality. Nevertheless, we would like to be able to explore a deeper notion of explanatory relevance relation for explanations and it may be interesting to investigate its connections with relevant logics too.

We would like to study the effects of modulating the kind of relevance between explananda and *explanans*. It is plausible that the kind of relevance that is adequate varies from context to context. Besides, it is likely that different default or paraconsistent logics[19] may provide different resources for the representation and analysis of different kinds of explanation.

Acknowledgements The authors would like to express their gratitude to Dr. María Alicia Pazos for her valuable discussions of previous versions of this work. Her comments enriched the final result. This research also benefited from the discussions carried out in the Special Interest Group in Non-Monotonicity, *SIGNO-MON*, and from the group of Formal Epistemology and Non Classical Logics of the Research Program in Science and Philosophy, *PRINCIPHIA*.

References

1. Achinstein, P. 1985. *The Nature of Explanation*. Oxford University Press.
2. Carnap, Rudolf. 1950. *Logical Foundations of Probability*. Chicago: University of Chicago Press.
3. Carnielli, Walter, and Marcelo Coniglio. 2016. *Paraconsistent Logic Consistency, Contradiction and Negation*. Logic, Epistemology and the Unity of Science 40. USA: Springer
4. Da Costa N.C.A. 1974. On the theory of Inconsistent Formal Systems. *Notre Dame Journal of Formal Logic*, 15-4.
5. D'Ottaviano, Itala M. L., and H. Feitosa. 2000. Paraconsistent Logics and Translations. *Synthese*, 125-1/2.
6. Gaytán Cabrera, David. 2014. Un modelo no monotónico y paraconsistente de explicación científica. PhD thesis, UNAM.
7. Hempel, Carl, G., and Paul Oppenheim. 1948. Studies in the Logic of Explanation. *Philosophy of Science*, 15.
8. Hempel, Carl G. 1965. *Aspects of Scientific Explanation and Other Essays in the Philosophy of Science*. New York: Free Press and Collier Macmillan.
9. Hempel, Carl G. 1988. Provisoes: A Problem Concerning the Inferential Function of Scientific Theories. *Erkenntnis*, 28.
10. Lewis, David. 1986. *Philosophical Papers II*. Oxford: Oxford University Press.
11. Maher, Patrick. 2007. Explication defended. *Studia Logica*, 86-2.
12. Marek, V. Wiktor, and Miroslaw Truszczynski. 1991. *Non Monotonic Logic Context-Dependent Reasoning*. USA: Springer
13. Morado, Raymundo. 2004. Problemas filosóficos de la Lógica No Monotónica. In Raúl Orayen, Alberto Moretti, Enciclopedia Iberoamericana de Filosofía, **27**, Filosofía de la Lógica, Trotta.
14. Nowak, Stefan. 1977. *Methodology of Sociological Research*. Poland: Reidel Publishing Company.

[19]Variations that could be very fruitful to explore are the Logics of Formal Inconsistency, as they are developed in [3].

15. Orayen, Raúl. 1988. Acerca de la adecuación de los modelos formales de la explicación científica. In Antologa de la lógica en América Latina, C.E.L.I.J.S. Universidad de Carabobo, Valencia, Venezuela.
16. Railton, Peter. 1978. A Deductive-Nomological Model of Probabilistic Explanation. *Philosophy of Science*, 45: 206–226.
17. Reiter, Raymond. 1980. A Logic for Default Reasoning. *Artificial Intelligence*, 13: 81–123.
18. Salmon, W.C. 1984. *Four Decades of Scientific Explanation*. University of Pittsburgh.
19. Salmon, W.C. 1984. *Scientific Explanation and the Causal Structure of the World*. Princeton University Press.
20. Van Fraassen, Bas. 1980. *The Scientific Image*. Oxford University Press.

Para-Disagreement Logics and Their Implementation Through Embedding in Coq and SMT

Bruno Woltzenlogel Paleo

1 Introduction

On closer inspection many apparent contradictions turn out to be mere disagreements between distinct sources of information. For example, if a source s_1 says P and a source s_2 says $\neg P$, their disagreement would only become an actual contradiction if we naively merged what they say into our own knowledge base. In this case, our own knowledge base would entail $P \wedge \neg P$ and would, therefore, be inconsistent. Although we could use traditional paraconsistent logics to avoid this kind of inconsistency's worst consequences, this would be an unsatisfactory approach, because the inconsistency in this case was clearly just a result of our indiscriminate use of knowledge originating from distinct mutually contradictory sources.

This paper proposes a new logical paradigm through which disagreements can be expressed and resolved. A *possible worlds* semantics is used (cf. Sect. 3), and each source denotes a world. Logical sentences of the form $@_s P$ express that source s claims P and denote that P is true at the world denoted by s. Within these logics, we can merge conflicting information more cautiously. For instance, our knowledge base would entail $(@_{s_1} P) \wedge (@_{s_2} \neg P)$ and, as desired, no inconsistency would follow from the disagreement between s_1 and s_2 with respect to P.

Section 4 explores a few different behaviours, attitudes and procedures that people typically use to resolve disagreements and to aggregate their opinions and beliefs in order to reach common collective decisions in the social groups to which they belong. They include *consensus*, *dictatorship*, *trust* and *voting*. All these disagreement resolution methods can be formalized within the proposed logical paradigm, and it may be considered that each method leads to a different logic conforming to the proposed paradigm. As discussed in Sect. 4.4, the voting method requires

B. Woltzenlogel Paleo (✉)
Canberra, ACT, Australia
e-mail: bruno.wp@gmail.com

W. Carnielli and J. Malinowski (eds.), *Contradictions, from Consistency to Inconsistency*, Trends in Logic 47, https://doi.org/10.1007/978-3-319-98797-2_7

formulating the possible worlds semantics in a way that is, from a technical perspective, subtly different from traditional formulations (cf. Sect. 3).

One of the main goals during the development of para-disagreement logics, besides simplicity and conceptual adequacy, was ease of implementation of automated reasoning tools for these logics. To show that this goal has been achieved, Sects. 5 and 6 exhibit and discuss semantical embeddings of the logics into, respectively, the non-extensional higher-order logic of the interactive proof assistant `Coq` and a fragment of first-order logic extended with datatypes and linear integer arithmetic supported by the automatic SMT-solver `Z3`.

2 Syntax

As shown in Definition 1, the basic language of our logics is the usual language of propositional logic extended with the usual box and diamond operators of modal logics [12], and the @ operator and the (here explicit) *in* nominal operator from hybrid logics [2]. This basic syntax is extended in Sect. 4.

Definition 1 Given countably infinite sets $\mathscr{P}$ and $\mathscr{S}$ of, respectively, propositional symbols and information sources, the set of formulas *formulas* $\mathscr{L}$ is the smallest set satisfying:

- if $p \in \mathscr{P}$, then $p \in \mathscr{L}$.
- if $\varphi \in \mathscr{L}$, then $\neg\varphi \in \mathscr{L}$.
- if $\varphi_1 \in \mathscr{L}$ and $\varphi_2 \in \mathscr{L}$, then:
  - $\varphi_1 \wedge \varphi_2 \in \mathscr{L}$.
  - $\varphi_1 \vee \varphi_2 \in \mathscr{L}$.
  - $\varphi_1 \rightarrow \varphi_2 \in \mathscr{L}$.
- if $\varphi \in \mathscr{L}$, then:
  - $\Box\varphi \in \mathscr{L}$.
  - $\Diamond\varphi \in \mathscr{L}$.
- if $s \in \mathscr{S}$ and $\varphi \in \mathscr{L}$, then $@_s\varphi \in \mathscr{L}$.
- if $s \in \mathscr{S}$, then $in(s) \in \mathscr{L}$.
- if $s \in \mathscr{S}$ and $g \in \mathscr{S}$, then $(s \in g) \in \mathscr{L}$.

3 Semantics

The semantics for para-disagreement logics, as described in Definitions 2 and 3, is conceptually similar to typical possible worlds semantics for modal logics, essentially differing only in the representation of world reachability.

Definition 2 A *model* is a tuple $M := \langle W, R, I_{\mathscr{S}}, I_{\mathscr{P}} \rangle$, where W is a countable set of *worlds*, $R : W \longrightarrow List[W]$ is a function that maps each world to the list of worlds reachable from it, $I_{\mathscr{S}} : \mathscr{S} \longrightarrow W$ is an interpretation function that maps each source to the world it denotes, and $I_{\mathscr{P}} : W \times \mathscr{P} \longrightarrow \{\top, \bot\}$ is an interpretation function that maps each world and atomic proposition to a truth value.

Definition 3 The *truth* of a formula φ in a world w of a model $M := \langle W, R, I_{\mathscr{S}}, I_{\mathscr{P}} \rangle$ is denoted $M \vDash_w \varphi$ and is defined recursively as follows:

- $M \vDash_w p$ for every $p \in \mathscr{P}$ such that $I_{\mathscr{P}}(w, p) = \top$
- $M \vDash_w \neg\varphi$ iff $M \nvDash_w \varphi$.
- $M \vDash_w \varphi_1 \wedge \varphi_2$ iff $M \vDash_w \varphi_1$ and $M \vDash_w \varphi_2$
- $M \vDash_w \varphi_1 \vee \varphi_2$ iff $M \vDash_w \varphi_1$ or $M \vDash_w \varphi_2$
- $M \vDash_w \varphi_1 \rightarrow \varphi_2$ iff $M \nvDash_w \varphi_1$ or $M \vDash_w \varphi_2$
- $M \vDash_w \Box\varphi$ iff $M \vDash_{w'} \varphi$ for every $w' \in R(w)$
- $M \vDash_w \Diamond\varphi$ iff $M \vDash_{w'} \varphi$ for some $w' \in R(w)$
- $M \vDash_w @_s\varphi$ iff $M \vDash_{I_{\mathscr{S}}(s)} \varphi$
- $M \vDash_w in(s)$ iff $I_{\mathscr{S}}(s) = w$
- $M \vDash_w s \in g$ iff $I_{\mathscr{S}}(s) \in R(I_{\mathscr{S}}(g))$

Definition 4 A set of logical sentences $\varphi_1, \ldots, \varphi_n$ *entails* another logical sentence φ, denoted $\varphi_1, \ldots, \varphi_n \vDash \varphi$, iff $M \vDash_w \varphi$ for all w and for every model M such that $M \vDash_w \varphi_i$ for $1 \leq i \leq n$ and for all w.

The main technical difference between the semantics described above and the usual possible worlds semantics is that reachability of worlds is represented not as a binary relation between worlds, but as a function that maps a world to a list of its reachable worlds. Consequently, in the models used in para-disagreement logics, the number of worlds reachable from any world is always finite (because lists are finite), whereas this number could be countably infinite when reachability is represented as a binary relation between worlds. Having only a finite number of reachable worlds is important for the disagreement resolution method described in Sect. 4.4.

Definition 5 A source s is a *group* in a model $M := \langle W, R, I_{\mathscr{S}}, I_{\mathscr{P}} \rangle$ iff $R(I_{\mathscr{S}}(w))$ is non-empty (i.e. the world denoted by s has reachable worlds). Any source s'

In para-disagreement logics, the formula $@_s\varphi$ can be read as "*s claims* φ" or "*s is of the opinion that* φ". The formula $s \in g$ can be read as "*s* is a *participant* in the *group* *g*". Para-disagreement logics are concerned with establishing relationships between the claims of a group and the claims of its participants (or even non-participants).

The notion of group does not need to be understood in a narrow sense. From a broader and abstract perspective, for example, a person whose opinions are strongly influenced by three newspapers could be modelled as a "group" where those media sources are participants; and a software system that combines data from several different databases could be modelled as a "group" that has those databases as participants.

4 Methods for Information Aggregation and Disagreement Resolution

The main concern from this point on is to investigate possible relationships between the opinions of a group and the opinions of its members. In general, the group needs to aggregate the opinions of its members and resolve disagreements. There are potentially many methods to do that, and para-disagreement logics do not aim to advocate in favour of one method, but rather to be flexible and expressive enough to allow a wide variety of methods to be formalized and possibly combined to reason simultaneously about groups with distinct disagreement resolution behaviours. The next subsections discuss a few.

4.1 Consensus

One of the simplest methods for a group to aggregate information from its members is *consensus*: if all members of a group g claim something, then g claims it as well. In logical form, the consensus axiom schema for a fixed[1] group g is:

$$@_g(\Box\varphi \rightarrow \varphi)$$

More strongly, a para-disagreement logic assuming the more general and well-known T axiom schema ($\Box\varphi \rightarrow \varphi$) would be stating that all groups respect consensus.

The obvious limitation of consensus is that it says nothing about the opinions of a group on matters on which its members disagree.

4.2 Dictatorship

Dictatorship is perhaps the simplest method to aggregate information in cases when there is disagreement. The opinion of the group is simply dictated by a distinguished source. The dictatorship axiom schema for a fixed group g with a fixed dictator d is[2]:

$$@_g(@_d\varphi \rightarrow \varphi)$$

Furthermore, to show, for instance, that the dictator d has dissidents in his dictatorship g, it suffices to find a proposition q for which $@_g(@_d q \wedge \Diamond\neg q)$. This

[1] By “fixed“ it is meant that g is not a schema variable, but a constant. The consensus axiom schema does *not* hold for all g. It is the responsibility of a user of the logic to instantiate and assert the axiom schema for any particular g^* that is a consensus group.

[2] Note that d does not need to be a participant of g.

illustrates that it is easy to formalize dictatorships and reason about their situations in para-disagreement logics.

A serious problem of dictatorships is that they blindly follow the opinions of the dictator and simply ignore the opinions of other members of the group, who might have greater expertise on certain topics than the dictator.

4.3 Expertise-Restricted Trust

In human society, groups often follow restricted forms of dictatorship, in which the opinions and decisions of the group with respect to a proposition pertaining to a certain topic is dictated by a member of the group with declared expertise in that topic. For example, we are normally willing to trust the opinions of doctors on matters related to our health and the opinions of lawyers on legal matters, but not conversely; and a company's technological decisions are ultimately taken by the CTO, whereas its financial decisions by the CFO.

In order to be able to formalize groups with information aggregation based on expertise-restricted trust, it is necessary to extend the basic syntax of para-disagreement logic. One possibility is to extend it with the notion of topic and special operators to talk about expertise on topics and pertinence to topics. For example, the fact that an information source s has expertise on topic t could be expressed by the formula $E(s, t)$ and the fact that a proposition p is about a topic t could be expressed by the formula $A(p, t)$.[3] Assuming, for the sake of simplicity, that experts on a given topic do not disagree on propositions related to that topic, the axiom schema stating that g trusts its experts can be written as:

$$@_g((E(\varepsilon, \tau) \wedge A(\varphi, \tau)) \rightarrow (@_\varepsilon \varphi \rightarrow \varphi))$$

Despite the additional operators, the logic is still quantifier-free, and hence essentially propositional.

4.4 Voting

Another common method to manage disagreements, especially in democratic societies, is to vote. There exist various vote-counting methods [23] and, in principle, para-disagreement logics could be extended with any vote-counting method. But *majority voting* is one of the simplest, and it is already sufficiently interesting to illustrate a vote-based para-disagreement logic.

---

[3] A is assumed to be an intensional operator: the truth values of $A(p, t)$ and $A(q, t)$ may differ from each other even when $p \leftrightarrow q$.

The syntax of the logic needs to be extended with an intensional *vote* operator, a *majority* modality and an *aggregation* connective, as shown below.

Definition 6 Given countably infinite sets $\mathscr{P}$ and $\mathscr{S}$ of, respectively, propositional symbols and information sources, the set of formulas *formulas* $\mathscr{L}_V$ is the smallest set satisfying the clauses in Definition 1 (with $\mathscr{L}$ replaced by $\mathscr{L}_V$) and also:

- if $\varphi \in \mathscr{L}_V$, then $\mathrm{V}(\varphi) \in \mathscr{L}_V$.
- if $\varphi \in \mathscr{L}_V$, then $\mathfrak{M}\varphi \in \mathscr{L}_V$.
- if $g, s_1, \ldots, s_n \in \mathscr{S}$, then $g \prec [s_1, \ldots, s_n] \in \mathscr{L}_V$.

The intended readings of these new kinds of formulas are:

- $\mathrm{V}(\varphi)$: φ is a proposition at issue in the vote.
- $\mathfrak{M}\varphi$: the majority chooses φ (over $\neg\varphi$).
- $g \prec [s_1, \ldots, s_n]$: the group g aggregates the sources s_1, ..., s_n.

The semantics also has to be extended to cope with new syntactical constructions. An interpretation for the vote operator is needed, as shown in Definition 7, and the notion of truth needs to be extended to formulas containing the vote operator, the majority modality and the aggregation connective, as stated in Definition 8.

Definition 7 A *model* is a tuple $M := \langle W, R, I_{\mathscr{S}}, I_{\mathscr{P}}, I_V \rangle$, where W, R, $I_{\mathscr{S}}$ and $I_{\mathscr{P}}$ are as in Definition 2 and $I_V : W \times \mathscr{L}_V \longrightarrow \{\top, \bot\}$ is an interpretation function that maps each world and formula to a truth value.

Definition 8 The *truth* of a formula φ in a world w of a model $M := \langle W, R, I_{\mathscr{S}}, I_{\mathscr{P}} \rangle$ is denoted $M \vDash_w \varphi$ and is defined recursively using the cases shown in Definition 3 as well as the following:

- $M \vDash_w \mathrm{V}(\varphi)$ iff $I_V(w, \varphi) = \top$
- $M \vDash_w \mathfrak{M}\varphi$ iff there are more $w' \in R(w)$ such that $M \vDash_{w'} \varphi$ than $w'' \in R(w)$ such that $M \vDash_{w''} \neg\varphi$
- $M \vDash_w g \prec [s_1, \ldots, s_n]$ iff $R(I_{\mathscr{S}}(g)) = [I_{\mathscr{S}}(s_1), \ldots, I_{\mathscr{S}}(s_n)]$

It is now clearer why a technically different representation of reachability of worlds (using a function mapping worlds to (finite) lists of worlds, instead of a binary relation on worlds) was chosen. The definition of truth for a formula starting with a majority modality requires comparing the number of worlds w' for which φ is true with the number of worlds w'' for which $\neg\varphi$ is true. By requiring every world to have finitely many reachable worlds, the unclear corner case of comparing infinite numbers is simply avoided. The reason for choosing lists, in particular, is discussed in Sect. 5.

The majority voting axiom schema, stating that an arbitrary proposition is claimed by a fixed group g if the majority of its members asserts it and it is voted, can then be expressed as:

$$@_g((\mathfrak{M}\varphi \wedge \mathrm{V}(\varphi)) \to \varphi)$$

It is important to note that the vote operator is *intensional* with respect to its argument, in the following sense: the truth value of $\mathrm{V}(p)$ depends on p's intension (i.e. the proposition itself) and not only on p's extension (i.e. its denotation, its truth value). Consequently, $\mathrm{V}(p)$ and $\mathrm{V}(q)$ may have different truth values, even if p and q have the same truth values. Furthermore, if a group votes on a proposition p, this does not mean that the group omnisciently votes on all other propositions that it considers equivalent to p.

At this point one may wonder why the vote operator is needed. The following example, which is loosely inspired by the real-world case of the Brazilian president's impeachment in 2016, shows that a modified majority voting axiom schema without the vote guard (i.e. $@_g(\mathfrak{M}p \rightarrow p)$) would not be very useful, because then the opinions of a group would often be contradictory.

	C ?	$C \rightarrow I$?	I ?
Senator a	C	$C \rightarrow I$	I
Senator b	$\neg C$	$C \rightarrow I$	$\neg I$
Senator c	C	$\neg(C \rightarrow I)$	$\neg I$

Example 1 Suppose that a senate g composed of three senators a, b and c (i.e. $g \prec [a, b, c]$) has to decide whether the president should be impeached (proposition I). To come to that decision, the senators must deliberate on whether the president has committed a certain crime (proposition C) and whether that crime is a sufficient reason for impeachment ($C \rightarrow I$). Suppose that the senators think according to the following table:

Let S=$\{@_a(C \wedge (C \rightarrow I) \wedge I), @_b(\neg C \wedge (C \rightarrow I) \wedge \neg I), @_c(C \wedge \neg(C \rightarrow I) \wedge \neg I), g \prec [a, b, c]\}$. Then $S \vDash @_g(\mathfrak{M}C \wedge \mathfrak{M}(C \rightarrow I) \wedge \mathfrak{M}\neg I)$. If $@_g(\mathfrak{M}p \rightarrow p)$ is admitted, then:

- $S, @_g(\mathfrak{M}p \rightarrow p) \vDash @_g\neg I$, because $S \vDash @_g(\mathfrak{M}\neg I)$.
- $S, @_g(\mathfrak{M}p \rightarrow p) \vDash @_g I$, because $S \vDash @_g(\mathfrak{M}C)$, and thus $S, @_g(\mathfrak{M}p \rightarrow p) \vDash @_g C$, and $S \vDash @_g(\mathfrak{M}(C \rightarrow I)$, and thus $S, @_g(\mathfrak{M}p \rightarrow p) \vDash @_g(C \rightarrow I)$.

And hence g's opinions are contradictory (under S and $@_g(\mathfrak{M}p \rightarrow p)$) because both a proposition and its negation must be true. To avoid this, the senate must use the guarded majority voting axiom schema and decide a priori on which propositions it is going to vote. It may choose to vote on the conclusion (I) or on the premises (C and $C \rightarrow I$), but it shouldn't vote simultaneously on both. In either case, it is possible to reason about the outcome of the senate's decision process:

- in the first case (voting directly on the conclusion), $S, @_g((\mathfrak{M}p \wedge \mathrm{V}(p)) \rightarrow p), \mathrm{V}(I) \vDash @_g\neg I$ and hence the senate will decide not to impeach the president.
- in the second case (voting on the premises, and then deciding on the impeachment by logical reasoning), $S, @_g((\mathfrak{M}p \wedge \mathrm{V}(p)) \rightarrow p), \mathrm{V}(C), \mathrm{V}(C \rightarrow I) \vDash @_g I$ and hence the senate will decide to impeach the president.

One could also question the inclusion of the aggregation connective in the language and wonder if it would not be possible to use the diamond $\Diamond$ and the explicit nominal *in* operators instead, since together they can also state that a world is reachable from another. The following example discusses this.

Example 2 Consider the same scenario from Example 1 above, but let S' be S with the statement $g \prec [a, b, c]$ replaced by $@_g(\Diamond in(a) \wedge \Diamond in(b) \wedge \Diamond in(c))$. Superficially S' and S may appear equivalent, but in fact they are not. S' admits models where the world denoted by g has other reachable worlds besides those denoted by a, b and c, whereas in S the worlds reachable from the world denoted by g are exactly only those denoted by a, b and c. Therefore, $g \prec [a, b, c]$ is actually a stronger statement than $@_g(\Diamond in(a) \wedge \Diamond in(b) \wedge \Diamond in(c))$. A consequence of this fact is that, whereas with the former statement S, $@_g((\mathfrak{M}p \wedge \mathrm{V}(p)) \rightarrow p), \mathrm{V}(I) \models @_g \neg I$, with the latter statement S', $@_g((\mathfrak{M}p \wedge \mathrm{V}(p)) \rightarrow p), \mathrm{V}(I) \not\models @_g \neg I$, because S' gives only partial information about the worlds reachable from the world denoted by g. There is a model M of S', $@_g((\mathfrak{M}p \wedge \mathrm{V}(p)) \rightarrow p), \mathrm{V}(I)$ where the world denoted by g has several other reachable worlds where I is true and then $M \models @_g I$.

Instead of extending the language with the aggregation connective, the language could have been extended with quantification over sources and equality of sources. In this case, a statement such as $g \prec [a, b, c]$ could be replaced by $@_g(\Diamond in(a) \wedge \Diamond in(b) \wedge \Diamond in(c) \wedge \forall x.(\Diamond in(x) \rightarrow x = a \vee x = b \vee x = c))$. However, not only the aggregation connective is simpler, more concise and more convenient, but it also eases the implementation of counting, as discussed in the next two sections.

5 Embedding of Para-Disagreement Logics in Coq

A tool to support formal reasoning within para-disagreement logics can be implemented through a shallow embedding of the semantics of para-disagreement logics in the `Coq` proof assistant [11], which is based on the calculus of inductive constructions [20] for a non-extensional type-theoretical higher-order logic. The first step is to declare a type for worlds and the reachability function:

```
Parameter W: Type. (* Type for worlds *)
Parameter r: W -> list W. (* Reachability function *)
```

Then the type of propositions of para-disagreement logics is defined as the function type of functions that take a world and return a `Coq` proposition (i.e Prop):

```
Definition o := W -> Prop. (* Type of modal propositions *)
```

The propositional connectives operating on the lifted modal propositions are defined as functions taking modal propositions and a world, and returning a `Coq` proposition. Through currying (partial application of functions to arguments), such connectives can also be seen as taking modal propositions and returning a modal proposition (i.e. a function that takes a world and returns a `Coq` proposition). Notations are declared to allow the new lifted connectives to be written down exactly as `Coq`'s built-in connectives, but with an "m" prefix.

```
Definition mnot (p: o)(w: W) := ˜ (p w).
Notation "m˜  p" := (mnot p) (at level 74, right associativity).

Definition mand (p q:o)(w: W) := (p w) / (q w).
Notation "p m/ q" := (mand p q)
   (at level 79, right associativity).

Definition mor (p q:o)(w: W) := (p w)   (q w).
Notation "p m q" := (mor p q) (at level 79, right associativity).

Definition mimplies (p q:o)(w: W) := (p w) -> (q w).
Notation "p m-> q" := (mimplies p q)
   (at level 99, right associativity).

Definition mequiv (p q:o)(w: W) := (p w) <-> (q w).
Notation "p m<-> q" := (mequiv p q)
   (at level 99, right associativity).
```

The use of a reachability function mapping worlds to their reachable worlds instead of a binary reachability relation between worlds requires that the box and diamond modalities be defined in a different way, using an auxiliary function that traverses the list of reachable worlds.

```
(* auxiliary function to checks if an element is in a list. *)
Fixpoint is_in A: Type (x: A) (l: list A) := match l with
  | nil => False
  | (cons h tail) => x = h  (is_in x tail)
end.

(* Box Modal Operator *)
Definition box (p: o) :=
           fun w => forall w1, (is_in w1 (r w)) -> (p w1).

(* Diamond Modal Operator *)
Definition dia (p: o) :=
           fun w => exists w1, (is_in w1 (r w)) / (p w1).
```

The @ modality and the explicit nominal operator *in* borrowed from hybrid logics are defined as expected, and a notation is declared to allow the special symbol `@` to stand for the modality.

```
Definition At (w: W)(p: o) := fun w0: W => (p w).
Notation "'@' w p" := (At w p)
  (at level 200, w ident, right associativity) : type_scope.

Definition In (w: W) := fun w0 => w = w0.
```

For the sake of simplicity, it is assumed here that the set of source symbols $\mathscr{S}$ and the set of worlds W denoted by the sources coincide. In other words, the interpretation function $I_{\mathscr{S}}$ is assumed to be the identity function. In the `Coq` embedding, this is reflected in the usage of the type `W` not only for worlds but also for sources, as seen in the type of the @ modality.

The embedding of the basic language of para-disagreement logics in Coq is completed with the definition of the aggregation connective and its corresponding notation:

```
Definition aggregation (g: W)(l: list W): o :=
           fun w: W => (r g) = l.
Notation "g '<<' l" := (aggregation g l)
  (at level 70) : type_scope.
```

And finally, quotes are used as notation for truth of a modal proposition in all worlds:

```
Definition UniversallyTrue (p: o) := forall w, p w.
Notation "' p '" := (UniversallyTrue p).
Ltac mv := match goal with [|- (UniversallyTrue _)] => intro end.
```

With all the basic language ready, it is time to move on to the extensions described in Sect. 4.4. As the majority modality requires counting, an auxiliary count function is defined:

```
Parameter dec: forall (f: o)(w: W), f w + ~ (f w).

Fixpoint count (p: o) (l: list W) := match l with
  | [] => 0
  | head::tail => if (dec p head)
                  then (1 + (count p tail))
                  else (count p tail)
end.
```

The function count needs to traverse all the reachable worlds and count on how many of them the modal proposition is true. Lists are the simplest traversable and collection datatype, and that is why it was chosen here and also as the return type of the reachability function in Definition ??. The parameter dec is needed to conform with the typing requirements of the "if …then …else …" expression.

Once the count function is available, defining the majority modality can be easily done as follows:

```
Definition M (p: o) :=
           fun w: W => ((count p (r w)) > (count (m~ p) (r w))).
```

Next the intensional vote operator is declared, together with an axiom stating that it is invariant with respect to negation of its argument.

```
(* vote operator *)
Parameter V: o -> o.

Axiom vote_invariant_wrt_negation:
      'mforall p, (V p) m<-> (V (m~ p)) '.
```

And then finally the majority axiom schema can be declared:

```
Axiom majority_axiom: 'mforall p, ((V p) m-> ((M p) m-> p))'.
```

Now that the para-disagreement logic is fully embedded, the impeachment example can be formalized as shown below:

```
(* The three senators *)
Parameters a b c: W.

(* The senate containing the three senators *)
Parameter g: W.
Axiom e: '(g << [a; b ; c])'.


(* The two atomic propositions *)
(* proposition that the president committed a crime *)
Parameters C: o.
(* proposition that the president should be impeached *)
Parameters I: o.

(* The senators' opinions *)
Axiom a_claims_C: '(@ a C)'.
Axiom a_claims_C_implies_I: '@ a (C m-> I)'.
(* a's third opinion is not independent *)
Lemma a_claims_I: '@ a I ' .
Proof. mv.
apply (a_claims_C_implies_I w). apply (a_claims_C w).
Qed.

Axiom b_claims_not_C: '@ b (m~ C)'.
Axiom b_claims_C_implies_I: '@ b (C m-> I)'.
Axiom b_claims_not_I: '@ b (m~ I)'.

Axiom c_claims_C: '(@ c C)'.
Axiom c_claims_not_C_implies_I: '@ c (m~ (C m-> I))'.
Axiom c_claims_not_I: '(@ c (m~ I))'.

(* The propositions that have been voted in the senate *)
Axiom C_is_voted: '@ g (V C)'.
Axiom C_implies_I_is_voted: '@ g (V (C m-> I))'.
```

From the axioms stated above, it is now possible to prove that the majority claims that the president committed a crime:

```
Lemma majority_claims_C: '@ g (M C)'.
Proof.
mv.
unfold At; unfold M.
rewrite (e w).
assert (C a); [apply (a_claims_C w) | auto].
assert ((m~ C) b); [apply (b_claims_not_C w) | auto].
assert (C c); [apply (c_claims_C w) | auto].
unfold count.
destruct (dec C a); [auto | contradiction].
destruct (dec C c); [auto | contradiction].
destruct (dec (m~ C) b); [auto | contradiction].
destruct (dec (m~ C) a); [contradiction | auto].
destruct (dec C b); [contradiction | auto].
destruct (dec (m~ C) c); [contradiction | auto].
Qed.
```

However, the proof above is tedious, requiring the user to interactively count the senators that (dis)agree with the claim. In order to automate the counting, a new tactic can be implemented using Coq's Ltac language, as shown below:

```
Ltac count db :=
  match goal with
  |- context [if dec ?q ?x then _ else _] =>
      destruct (dec q x);
      firstorder with db;
      count db
  end.
```

The count tactic receives a hint database db of axioms and tries to automatcally decide the conditions of if-then-else statements in the goal using the firstorder tactic with the given database. It is a recursive tactic that reapplies itself until it fails.

With shallow embeddings, it is also often the case that definitions need to be unfolded for a goal to be proven. To automate the unfolding, the unfold_pdl tactic defined below repeatedly tries to unfold all defined connectives, quantifiers and operators included in the modal unfold hint database occurring both in the conclusion and in the hypotheses of the goal.

```
Create HintDb modal.
Hint Unfold mimplies mequiv mnot mor mand
             dia box A E M At In UniversallyTrue count: modal.

Ltac unfold_pdl := try mv; repeat autounfold with modal;
                          repeat autounfold with modal in * |-.
```

Automation can be improved further with a tactic that combines the previously defined tactics with Coq's built-in auto and autorewrite tactics:

```
Ltac pdl_solve kb rb := unfold_pdl; autorewrite with rb;
                         try auto with kb modal; try count kb.
```

In the case of the impeachment example, the databases of facts and rewrite equalities can be created as shown below:

```
Create HintDb db.
Hint Resolve a b c g e C I: db.
Hint Resolve C_is_voted C_implies_I_is_voted: db.
Hint Resolve a_claims_C a_claims_C_implies_I a_claims_I: db.
Hint Resolve b_claims_not_C
             b_claims_C_implies_I
             b_claims_not_I: db.
Hint Resolve c_claims_C
             c_claims_not_C_implies_I
             c_claims_not_I: db.

Create HintDb rb.
Hint Rewrite e: rb.
```

The previous lemma can now be proved fully automatically:

```
Lemma majority_claims_C: '@ g (M C)'.
Proof.
pdl_solve db rb.
Qed.
```

And other lemmas can be proven fully automatically as well:

```
Lemma majority_claims_C_implies_I: '@ g (M (C m-> I))'.
Proof.
pdl_solve db rb.
Qed.

Lemma majority_claims_not_I: '@ g (M (m~ I))'.
Proof.
pdl_solve db rb.
Qed.
```

Full automation is not possible when the lemma to be proven depends on axioms and lemmas that have not been included in the hint databases. Furthermore `Coq`'s `auto` tactic ignores axioms and lemmas that have a universally quantified head, because such axioms and lemmas can match any goal and, therefore, the proof search may not terminate. The majority axiom schema has a universally quantified head, and that is why it has not been included in the hint database. Consequently, lemmas that depend on this axiom currently cannot be proven fully automatically. Nevertheless, in such cases, it often suffices to apply the majority axiom, include previously proved lemmas in the local context with `pose`, and solve the remaining goals automatically with the provided tactic.

```
Lemma g_claims_C: '(@ g C)'.
Proof.
pdl_solve db rb.
apply majority_axiom.
  pose C_is_voted; pdl_solve db rb.
  pose majority_claims_C; pdl_solve db rb.
Qed.
```

A less automatic but more efficient alternative to `pose` and `pdl_solve` is shown below for a similar lemma.

```
Lemma g_claims_C_implies_I: '@ g (C m-> I)'.
Proof.
pdl_solve db rb.
apply (majority_axiom g);
  [apply (C_implies_I_is_voted w) |
   apply (majority_claims_C_implies_I w)].
Qed.
```

And finally the president's impeachment can be shown:

```
Theorem g_claims_I: '(@ g I)'.
Proof.
pdl_solve db rb.
apply (g_claims_C_implies_I w).
exact (g_claims_C w).
Qed.
```

6 Embedding of Para-Disagreement Logics in SMT

To automate reasoning in para-disagreement logics even further, SMT-solvers (for satisfiability modulo theories) such as `Z3` [17], which are capable of dealing with lists, recursive function definitions and linear integer arithmetic are a natural choice.

The main difference to the previous embedding in `Coq` is that the standard language[4] of SMT-solvers is not a higher-order typed language but a first-order multi-sorted language. This already causes difficulty when declaring the sort/type for propositions. In constrast to the embedding in `Coq`, where propositions had a defined function type from worlds to `Coq`'s built-in `Prop`, function types/sorts are not available in the first-order multi-sorted first-order language of SMT-solvers. Therefore, propositions are assumed to be of a primitive (declared but undefined) sort of arity 0.

```
(declare-sort o 0) ;; sort for propositions
(declare-sort W 0) ;; sort for worlds
```

As in the `Coq` embedding, reachability is declared as a function from worlds to lists of worlds:

```
(declare-fun r (W) (List W))
```

As a consequence of the fact that the sort `o` of propositions is now a primitive sort and not a function type from worlds to the meta-logic's propositions, the connectives and modal operators cannot be simply defined as functions, as they were in the `Coq` embedding. Instead, they must first be declared without definition:

```
(declare-fun mnot (o) o)
(declare-fun mimp (o o) o)
(declare-fun mand (o o) o)
(declare-fun mor (o o) o)
(declare-fun box (o) o)
(declare-fun dia (o) o)
(declare-fun M (o) o)
(declare-fun vote (o) o)
```

And then their intended meanings have to be axiomatized with the help of a truth predicate:

```
(declare-fun T (o W) Bool) ;; (T p w) = "p is true at world w"

(assert (forall ((w W) (p o))
            (iff (T (mnot p) w) (not (T p w))   ) ))

(assert (forall ((w W) (p o) (q o))
            (iff (T (mimp p q) w) (=> (T p w) (T q w)) ) ))

(assert (forall ((w W) (p o) (q o))
            (iff (T (mand p q) w) (and (T p w) (T q w)) ) ))

(assert (forall ((w W) (p o) (q o))
            (iff (T (mor p q) w) (or (T p w) (T q w)) ) ))
```

The axiomatizations of box, diamond and majority make use of auxiliary recursive function definitions that traverse the list of reachable worlds and output a formula that checks, respectively, whether the given proposition is true in all, in at least one, and in the majority of reachable worlds:

```
;; Box
(define-fun-rec TInAll ((p o) (l (List W)) ) Bool
  (or (= l (as nil (List W)))
      (and (T p (head l)) (TInAll p (tail l)))     ))

(assert (forall ((w W) (p o)) (iff (T (box p) w)
                                   (TInAll p (r w))    ) ))

;; Diamond
(define-fun-rec TInOne ((p o) (l (List W)) ) Bool
  (and (not (= l (as nil (List W))))
       (or (T p (head l)) (TInOne p (tail l)))     ))

(assert (forall ((w W) (p o)) (iff (T (dia p) w)
                                   (TInOne p (r w))    ) ))

;; Majority Modality
(define-fun-rec count ((p o) (l (List W)) ) Int
  (ite (= l (as nil (List W))) 0
       (ite (T p (head l)) (+ 1 (count p (tail l)))
                           (count p (tail l))          )))

(assert (forall ((w W) (p o))(iff (T (M p) w)
                                 (> (count p (r w))
                                    (count (mnot p) (r w))))))
```

For convenience, a *validity* predicate is defined, analogously to the quotes in the `Coq` embedding.

```
(define-fun V ((p o)) Bool (forall ( (w W) ) (T p w)) )
```

And finally the majority axiom schema can be asserted:

```
(assert (forall ((p o)) (V (mimp (mand (vote p) (M p)) p) ) ))
```

Now that the para-disagreement logic has been embedded, the impeachment example can be formalized. First the senators and the senate itself are declared as worlds:

```
(declare-fun senA () W)
(declare-fun senB () W)
(declare-fun senC () W)
(declare-fun senate () W)
```

Then the fact that the senate aggregates the three senators is asserted.

```
(assert (= (r senate)
           (insert senA
           (insert senB
           (insert senC (as nil (List W)))))  ))
```

And the atomic propositions are declared and the senators' opinions can be asserted:

```
(declare-fun c () o) ;; the president committed a crime
(declare-fun i () o) ;; the president should suffer impeachment

(assert (T c senA))
(assert (T (mimp c i) senA))
(assert (T i senA))

(assert (T (mnot c) senB))
(assert (T (mimp c i) senB))
(assert (T (mnot i) senB))

(assert (T c senC))
(assert (T (mnot (mimp c i)) senC))
(assert (T (mnot i) senC))
```

The facts that the senators decided to vote on whether the president committed a crime and on whether the crime should imply impeachment are asserted.

```
(assert (T (vote c) senate))
(assert (T (vote (mimp c i)) senate))
```

To ask the SMT-solver whether the previous assertions entail impeachment, the negation of the impeachment conjecture should be asserted, as shown below. This is so because SMT-solvers are refutational theorem provers, proving conjectures by contradiction.

```
(assert (not (T i senate)) )
```

Finally, the commands `check-sat` and `get-proof` should be invoked.

```
(check-sat)
(get-proof)
```

Unfortunately, `Z3` fails to prove this conjecture, probably because SMT-solvers have incomplete quantifier instantiation heuristics, which may be failing to find the correct instantiation when there are defined connectives and operators. Fortunately, `Z3` can solve the problem with just a slight help. If the following unfolding of the majority axiom schema is manually asserted, `Z3` is able to do all the remaining logical and arithmetical reasoning fully automatically, and outputs a proof that is 35842 characters long.

```
(assert (forall ((p o) (w W))
   (=> (and (T (vote p) w) (T (M p) w)) (T p w) )))
```

7 Related Work

The idea of using modal logics to handle (apparent) contradictions can be traced back at least to Jaskowski's *discussive logics* [16]. However, the para-disagreement logics proposed here use the @ modality, thereby overcoming well-known issues [19] faced by Jaskowki due to his use of the $\Diamond$ modality instead. As in Jaskowki's logics, the $\Box$ modality acts like a consensus operator. $\Box P$ expresses that everybody claims P. Together with the T axiom ($\Box P \rightarrow P$), the behavior of $\Box$ is reminiscent of the $\circ$ consistency operator of *logics of formal inconsistency* with principles of gentle explosion [13].

Preferential and *Distance-based* paraconsistent logics [3] form an interesting class of logic that handles inconsistencies by considering most preferred or least distant valuations of a theory in order to determine the logical consequences of the theory. Although the use of preferential and numerical approaches may suggest a similarity with the voting-based para-disagreement logic presented in Sect. 4.4 or with para-disagreement logics where an information source is preferred (e.g. as in Sect. 4.3), the similarities are superficial and the logics are actually very different, simply because para-disagreement logics are *not* paraconsistent logics. Para-disagreement logics are classical, monotonic, modal logics, where the principle of explosion holds. Their goal is to deal with consistent theories containing formulas such as $@_{s_1} P \wedge @_{s_2} \neg P$, which express a disagreement between s_1 and s_2. In paraconsistent logics (including preferential and distance-based), on the other hand, the concern is to avoid the principle of explosion in the presence of inconsistencies in theories with formulas such as $P \wedge \neg P$. Another difference between para-disagreement logics and preferential or distance-based paraconsistent logics is that the latter's preferential or distance-based mechanisms for avoiding explosion in the presence of inconsistencies is extra-logical and rigidly built-in as part of the semantics, whereas the former's disagreement resolution mechanisms are expressible syntactically in the logic itself through axiom schemata that can be flexibly modified and even combined to suit various domains of applications, as not all disagreements ought to be resolved in the same way.

As (apparent) contradictions can be common for AI agents and databases handling data from various sources or from different points in time, it is not surprising that many tasks, such as *belief revision* [1], *information/data integration* [18], *database repair* [10] and *consistent query answering* [14, 21], share an interest with para-disagreement logics on the topic of tackling (apparent) contradictions. One important distinguishing characteristic of the framework of para-disagreement logics is that it does not advocate for a specific way of handling apparent contradictions but rather provides an expressive language that allows disagreements to be explicitly modeled and allows a wide variety of disagreement resolution mechanisms to be asserted through axiom schemata, as non-exhaustively exemplified in Sect. 4. Due to this generality, it may be the case that some concrete approaches proposed for belief revision, database repair, information integration and consistent query answering could be defined as concrete para-disagreement logics within the framework described here. In such cases, the para-disagreement logic framework would serve as an alternative classical modal foundation to define these approaches, which are often described from a non-classical, non-monotonic and paraconsistent standpoint. Despite its generality, however, certainly not all database-related approaches are amenable to be described as para-disagreement logics. An essential requirement in the para-disagreement logic framework is the ability to distinguish and name the sources of contradictory information. In a practical database setting, this requirement is not always satisfied. For instance: a single source may already contain contradictory information; or, maybe, even though the contradictory information originates from different sources or time points, it is not known anymore from which source or time point each piece of information originated. Database-related techniques that target such situations are clearly outside the scope of the para-disagreement logic framework.

The embedding of para-disagreement logics in `Coq` follows an approach previously used in the embedding of the modal logics **K**, **KB** and **S5** in `Coq` [8, 9] and related to the embedding of the same logics in `Isabelle` [5, 7] and TPTP THF [6]. However, that approach had to be modified (as explained in the previous sections), because para-disagreement logics require a different technical encoding of reachability between worlds, and it also had to be extended with arithmetical reasoning for counting worlds. Furthermore, the work presented here also discusses automation of reasoning in para-disagreement logic within `Coq`, whereas the previous work in [9] was concerned with interactive reasoning only.

Thanks to the maturity, efficiency and popularity of SAT-solvers, theorem provers (e.g. [15, 22]) for non-classical and modal logics have been implemented recently with architectures that use SAT-solvers as black-boxes. In contrast, the work presented here uses an SMT-solver. As the logics of SMT-solvers are more expressive than the classical propositional logic of SAT-solvers, non-classical and modal logics (even complex ones requiring arithmetical reasoning such as para-disagreement logics) can be fully embedded within the logics of SMT-solvers, and these solvers can then be used directly, with no need to build a separate prover having an SMT-solver as a black-box component.

8 Conclusion

The para-disagreement logics presented here constitute a new paradigm to deal with apparent contradictions that occur when different agents or sources of information have conflicting opinions about some propositions. Four different disagreement resolution methods were discussed, with special emphasis on a majority voting method. However, it is important to note that para-disagreement logics are a general framework that, in principle, can support other (possibly more sophisticated) disagreement resolution methods as well.

The development of para-disagreement logics required a formulation of possible worlds semantics that is technically different from the usual. Their embedding into the meta-logics of `Coq` and SMT-solvers also pushed further the state-of-the-art of the embedding approach, as it required the use of arithmetics, which was not necessary in previous work on simpler modal logics. At the same time, the successful (almost full) automation of para-disagreement logical reasoning within `Coq` and `Z3` attests the current level of maturity of these tools even for a domain of application for which they were not originally intended. And indeed, the embeddings described here expand the range of applications of classical interactive and automated theorem provers to the area of paraconsistent reasoning, broadly understood, at least when contradictions are merely apparent as a result of disagreement between clearly identifiable sources.

Although the focus here was on *propositional* para-disagreement logics, this was so just because the propositional level was sufficient to discuss the essence of para-disagreement logics. The embedding into the meta-logic of SMT-solvers could be easily extended to quantifier-free first-order logic, and the embedding into the meta-logic of `Coq` can be easily extended to rigid higher-order logic with constant or varying domains (i.e. with actualistic or possibilistic quantifiers).

As para-disagreement logics target *apparent* inconsistencies (e.g. disagreements such as $@_{s_1} P \wedge @_{s_2} \neg P$), they should be regarded as a complement, and not a replacement, to paraconsistent logics, which handle actual inconsistencies (e.g. $P \wedge \neg P$).

Acknowledgements The impeachment example used throughout this paper is inspired by a similar example presented by Viktor Wang in the ANU-Unicamp Paraconsistency Workshop. Der Author ist Stipendiat der Österreichischen Akademie der Wissenschaften (APART).

References

1. Alchourrón, C.E., P. Gärdenfors, and D. Makinson. 1985. On the logic of theory change: partial meet contraction and revision functions. *The Journal of Symbolic Logic* 50 (2): 510–530. https://doi.org/10.2307/2274239.
2. Areces, C., and P. Blackburn. 2010. Special issue on hybrid logics. *Journal of Applied Logic* 8(4): 303–304. https://doi.org/10.1016/j.jal.2010.10.001.
3. Arieli, O. 2008. Distance-based paraconsistent logics. *International Journal of Approximate Reasoning* 48(3): 766–783. https://doi.org/10.1016/j.ijar.2007.07.002.
4. Barrett, C., L.M. de Moura, S. Ranise, A. Stump, and C. Tinelli. 2010. The SMT-LIB initiative and the rise of SMT - (HVC 2010 award talk). In: S. Barner, I.G. Harris, D. Kroening, O. Raz (eds.) Hardware and Software: Verification and Testing - 6th International Haifa Verification Conference, HVC 2010, Haifa, Israel, October 4-7, 2010. Revised Selected Papers, *Lecture Notes in Computer Science*, vol. 6504, p. 3. Springer. https://doi.org/10.1007/978-3-642-19583-9_2.
5. Benzmüller, C., and B. Woltzenlogel Paleo. 2013. Gödel's god in isabelle/hol. Archive of Formal Proofs 2013. http://afp.sourceforge.net/entries/GoedelGod.shtml.
6. Benzmüller, C., and B. Woltzenlogel Paleo. 2014. Automating Gödel's ontological proof of god's existence with higher-order automated theorem provers. In *ECAI 2014 - 21st European Conference on Artificial Intelligence, 18-22 August 2014, Prague, Czech Republic - Including Prestigious Applications of Intelligent Systems (PAIS 2014)*, ed. by T. Schaub, G. Friedrich, B. O'Sullivan *Frontiers in Artificial Intelligence and Applications*, vol. 263, pp. 93–98. IOS Press. https://doi.org/10.3233/978-1-61499-419-0-93.
7. Benzmüller, C., B. Woltzenlogel Paleo. 2014. On logic embeddings and Gödel's god. In *Recent Trends in Algebraic Development Techniques*, *Theoretical Computer Science and General*

Issues, ed. by M. Codescu, R. Diaconescu, I. uu, vol. 9463. Springer. https://doi.org/10.1007/978-3-319-28114-8.
8. Benzmüller, C., and B. Woltzenlogel Paleo. 2015. Higher-order modal logics: Automation and applications. In *Reasoning Web. Web Logic Rules - 11th International Summer School 2015, Berlin, Germany, July 31 - August 4, 2015, Tutorial Lectures*, ed. by W. Faber, A. Paschke, *LNCS*, vol. 9203, pp. 32–74. Springer. https://doi.org/10.1007/978-3-319-21768-0_2.
9. Benzmüller, C., and B. Woltzenlogel Paleo. 2015. Interacting with modal logics in the Coq proof assistant. In *Computer Science - Theory and Applications - 10th International Computer Science Symposium in Russia, CSR 2015, Listvyanka, Russia, July 13–17, 2015, Proceedings*, pp. 398–411. https://doi.org/10.1007/978-3-319-20297-6_25.
10. Bertossi, L.E. 2011. *Database Repairing and Consistent Query Answering*. Synthesis Lectures on Data Management. Morgan & Claypool Publishers. https://doi.org/10.2200/S00379ED1V01Y201108DTM020.
11. Bertot, Y., and P. Casteran. 2004. *Interactive Theorem Proving and Program Development*. Springer.
12. Blackburn, P., M. de Rijke, and Y. Venema. 2001. *Modal Logic*. New York: Cambridge University Press.
13. Carnielli, W., M.E. Coniglio, and J. Marcos. 2007. *Logics of Formal Inconsistency*, pp. 1–93. Dordrecht: Springer Netherlands. https://doi.org/10.1007/978-1-4020-6324-4_1.
14. Chomicki, J. 2008. Consistent query answering: The first ten years. In *Scalable Uncertainty Management, Second International Conference, SUM 2008, Naples, Italy, October 1-3, 2008. Proceedings*, ed. by S. Greco, T. Lukasiewicz, *Lecture Notes in Computer Science*, vol. 5291, pp. 1–3. Springer. https://doi.org/10.1007/978-3-540-87993-0_1.
15. Claessen, K., and D. Rosén. 2015. SAT modulo intuitionistic implications. In *Logic for Programming, Artificial Intelligence, and Reasoning - 20th International Conference, LPAR-20 2015, Suva, Fiji, November 24-28, 2015, Proceedings*, ed. by M. Davis, A. Fehnker, A. McIver, A. Voronkov, *Lecture Notes in Computer Science*, vol. 9450, pp. 622–637. Springer. https://doi.org/10.1007/978-3-662-48899-7_43.
16. da Costa, N.C.A., and F.A. Doria. 1995. On jaskowski's discussive logic. *Studia Logica* 54(1): 33–60. https://doi.org/10.1007/BF01058531.
17. de Moura, L.M., and N. Bjørner. 2008. Z3: an efficient SMT solver. In *Tools and Algorithms for the Construction and Analysis of Systems, 14th International Conference, TACAS 2008, Held as Part of the Joint European Conferences on Theory and Practice of Software, ETAPS 2008, Budapest, Hungary, March 29-April 6, 2008. Proceedings*, ed. by C.R. Ramakrishnan, J. Rehof, *Lecture Notes in Computer Science*, vol. 4963, pp. 337–340. Springer. https://doi.org/10.1007/978-3-540-78800-3_24.
18. Lenzerini, M. 2002. Data integration: A theoretical perspective. In *Proceedings of the Twenty-first ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems, June 3-5, Madison, Wisconsin, USA*, ed. by L. Popa, S. Abiteboul, P.G. Kolaitis, pp. 233–246. ACM. https://doi.org/10.1145/543613.543644.
19. Marcos, J. 2005. Modality and paraconsistency. *The Logica Yearbook* 2004: 213–222.
20. Paulin-Mohring, C. 2015. Introduction to the calculus of inductive constructions. In *All about Proofs, Proofs for All, Mathematical Logic and Foundations*, ed. by D. Delahaye, B. Woltzenlogel Paleo. London: College Publications.
21. Staworko, S., J. Chomicki, and J. Marcinkowski. 2012. Prioritized repairing and consistent query answering in relational databases. *Annals of Mathematics and Artificial Intelligence* 64(2–3): 209–246. https://doi.org/10.1007/s10472-012-9288-8.
22. Zohar, Y., and A. Zamansky. 2016. Gen2sat: An automated tool for deciding derivability in analytic pure sequent calculi. In *Automated Reasoning - 8th International Joint Conference, IJCAR 2016, Coimbra, Portugal, June 27 - July 2, 2016, Proceedings*, ed. by N. Olivetti, A. Tiwari, *Lecture Notes in Computer Science*, vol. 9706, pp. 487–495. Springer. https://doi.org/10.1007/978-3-319-40229-1_33.
23. Zwicker, W.S. 2016. Introduction to the theory of voting. In *Handbook of Computational Social Choice*, ed. by F. Brandt, V. Conitzer, U. Endriss, J. Lang, A.D. Procaccia, pp. 23–56. Cambridge University Press. https://doi.org/10.1017/CBO9781107446984.003.

Asymptotic Quasi-completeness and ZFC

Mirna Džamonja and Marco Panza

Abstract The axioms ZFC of first order set theory are one of the best and most widely accepted, if not perfect, foundations used in mathematics. Just as the axioms of first order Peano Arithmetic, ZFC axioms form a recursively enumerable list of axioms, and are, then, subject to Gödel's Incompleteness Theorems. Hence, if they are assumed to be consistent, they are necessarily incomplete. This can be witnessed by various concrete statements, including the celebrated Continuum Hypothesis CH. The independence results about the infinite cardinals are so abundant that it often appears that ZFC can basically prove very little about such cardinals. However, we put forward a thesis that ZFC is actually very powerful at some infinite cardinals, but not at all of them. We have to move away from the first few and to look at limits of uncountable cardinals, such as $\aleph_\omega$. Specifically, we work with singular cardinals (which are necessarily limits) and we illustrate that at such cardinals there is a very serious limit to independence and that many statements which are known to be independent on regular cardinals become provable or refutable by ZFC at singulars. In a certain sense, which we explain, the behavior of the set-theoretic universe is asymptotically determined at singular cardinals by the behavior that the universe assumes at the smaller regular cardinals. Foundationally, ZFC provides an

M. Džamonja
School of Mathematics, University of East Anglia, Norwich NR4 7TJ, UK
e-mail: M.Dzamonja@uea.ac.uk
URL: http://www.mth.uea.ac.uk/people/md.html

M. Panza (✉)
CNRS, IHPST (CNRS and Université Paris 1, Panthéon-Sorbonne), 13 Rue de Four, 75006 Paris, France
e-mail: marco.panza@univ-paris1.fr
URL: http://www.ihpst.cnrs.fr/membres/membres-permanents/panza-marco

M. Panza
Chapman University, Orange, CA, USA
e-mail: panza@chapman.edu

W. Carnielli and J. Malinowski (eds.), *Contradictions, from Consistency to Inconsistency*, Trends in Logic 47, https://doi.org/10.1007/978-3-319-98797-2_8

asymptotically univocal image of the universe of sets around the singular cardinals. We also give a philosophical view accounting for the relevance of these claims in a platonistic perspective which is different from traditional mathematical platonism.

1 Introduction

Singular cardinals have a fascinating history related to an infamous event in which one mathematician tried to discredit another and ended up being himself proved wrong. As Menachem Kojman states in his historical article on singular cardinals [28], 'Singular cardinals appeared on the mathematical world stage two years before they were defined'. In a public lecture at the Third International Congress of Mathematics in 1904, Julius König claimed to have proved that the continuum could not be well-ordered, therefore showing that Cantor's Continuum Hypothesis does not make sense, since this would entail that $2^{\aleph_0}$, the (putative) cardinal of the continuum, is not well defined. This was not very pleasant for Cantor, who was not alerted in advance and who was in the audience. However, shortly after, Felix Hausdorff found a mistake in König's reasoning, which was to have used an incorrect lemma of Bernstein, proved by induction on all $\aleph$s by an argument that did not work at an uncountable $\aleph$ that happens not to have an immediate predecessor. Such cardinals were then named 'singulars' by König in 1905.

This is not the way we define singular cardinals today. The difference between the modern definition and König's has important consequences that deserve an explanation.

Today, in ZFC, we regard cardinals as initial ordinals and the inherited order lets us talk about the *cofinality* of a cardinal κ, namely $\mathrm{cf}(\kappa)$. In ZF, following von Neumann, an ordinal is defined to be a transitive set well-ordered by the membership relation. It can then be seen that each ordinal is exactly the set of ordinals strictly smaller than it. A cardinal is an ordinal that is not bijective with any ordinal smaller than it. For example, ω is a cardinal since it is the smallest infinite ordinal. Notice that every ordinal α is bijective with exactly one cardinal, which we denote by '$|\alpha|$' and call 'the cardinality of α'.

As such, the definition of the cardinality of an ordinal does not require the Axiom of Choice, but this axiom is required to ensure that every set has a cardinality. More precisely, in ZF one can prove that every well-ordered set is bijective with exactly one cardinal, but not that any set can be well-ordered. This is what the Axiom of Choice states (in one of its equivalent forms).

A set B is cofinal in an ordered set $(A, \leq)$ if and only if B is a subset of A and for any x in A there is a y in B such that $x \leq y$. The cofinality $\mathrm{cf}(A)$ of an ordered set A is the smallest cardinality of a set cofinal in A. This definition applies in particular when A is a cardinal κ considered as an ordered set with the relation $\leq$ between ordinals, giving us the definition of $\mathrm{cf}(\kappa)$. Any cardinal is obviously cofinal in itself.

Hence, from the definition above it immediately follows that, for any cardinal κ, $\mathrm{cf}(\kappa) \leq \kappa$. Given this, we say that a cardinal κ is *regular* if and only if $\mathrm{cf}(\kappa) = \kappa$. We say, then, that κ is *singular* just if it is not regular. For example, $\aleph_\omega$ and $\aleph_{\omega+\omega}$ and $\aleph_{\omega_1}$ are each singular.

Already in ZF, it can be easily shown that the class of ordinals is well-ordered by the membership relation. It follows that every cardinal has an immediate successor. If the former is λ, we denote the latter by 'λ^+'. A cardinal which has an immediate predecessor is said to be a *successor cardinal*. If a cardinal is not a successor, then we say that it is a *limit cardinal*.

The example of $\aleph_0$ is enough to show that a limit cardinal can be regular. While analyzing König's attempted proof, Hausdorff proved, appealing to the Axiom of Choice, that all successor cardinals are regular. This makes, for example, $\aleph_n$ regular for any natural n. But this cannot be proved without the Axiom of Choice. In the absence of Choice it may even happen that all uncountable cardinals are singular. Indeed, Motik Gitik [18] has given a model of ZF where all uncountable cardinals are not only singular, but they all also have cofinality $\aleph_0$ (the model is described in some detail at the end of Sect. 3). In addition, even if the Axiom of Choice is admitted, we cannot prove, just within ZFC, that there are cardinals other than $\aleph_0$ and 0 which are both limit and regular, that is, *weakly inaccessible*, as any such putative cardinal is usually called (this follows from Gödel's incompleteness). On the other hand, it is easy to see that there are unboundedly many singular cardinals. For example, $\aleph_{\alpha+\omega}$ is singular for any ordinal α.

All this came about since the existence of singular cardinals destroyed an infamous argument put forward in 1904. But there is little justification in the qualification 'singular' that was obtained because of this incident, since there are unboundedly many such cardinals. This qualification is even quite unfortunate because, in fact, it is at these cardinals that set theory behaves better: it turns out that at them the incompleteness phenomena of ZFC are much less present (although not totally absent) than at regular cardinals. We shall illustrate this fact by various recent mathematical findings. These results show that many statements which are known to be independent at regular cardinals become provable or refutable by ZFC at singulars, and so indicate that the behavior of the set-theoretic universe is asymptotically determined at singular cardinals by its features at the smaller regular cardinals. We could say, then, that even though ZFC is provably incomplete, asymptotically, at singular cardinals, it becomes quasi-complete since the possible features of universes of ZFC are limited in number, relative to the size of the singular in question. These facts invite a philosophical reflection.

The paper is organized as follows: Mathematical results that illustrate the mentioned facts are expounded in Sects. 2 and 3. The former contains results that by now are classic in set theory and it is written in a self-contained style. The latter contains results of contemporary research and is meant to reinforce the illustration offered by the former. This section is not written in a self-contained style, and it would be out of the scope of this paper to write it in this way. Section 2 also contains a historical perspective. Finally, some philosophical remarks are made in Sect. 4.

2 Modern History of the Singular Cardinals

One of the most famous (or infamous, depending on the point of view) problems in set theory is that of proving or refuting the Continuum Hypothesis (CH) and its generalisation to all infinite cardinals (GCH).

Cantor recursively defined two hierarchies of infinite cardinals, the $\aleph$s and the $\beth$s, the first based on the successor operation and the second on the power set operation: $\aleph_0 = \beth_0 = \omega$, $\aleph_{\alpha+1} = \aleph_\alpha^+$, $\beth_{\alpha+1} = 2^{\beth_\alpha}$, and for δ a non-zero limit ordinal $\aleph_\delta = \sup_{\beta<\delta} \aleph_\beta$, $\beth_\delta = \sup_{\beta<\delta} \beth_\beta$ (here we are using the notation 'sup(A)' for a set A of cardinals to denote the first cardinal greater or equal to all cardinals in A). A simple way to state GCH is to claim that these two hierarchies are the same: $\aleph_\alpha = \beth_\alpha$, for any α. Another way, merely involving the first hierarchy, is to claim that for every α we have $2^{\aleph_\alpha} = \aleph_\alpha^+$. CH is the specific instance $\aleph_1 = \beth_1$ or $2^{\aleph_0} = \aleph_1$. Insofar as $\beth_1 = |\mathbb{R}|$, CH can be reformulated as the claim that any infinite subset of the set of the real numbers admits a bijection either with the set of natural numbers or with the set of real numbers.

It is well known that, frustratingly, Cantor spent at least thirty years trying to prove CH. Hilbert choose the problem of proving or disproving GCH as the first item on his list of problems presented to the International Congress of Mathematics in 1900. In 1963 [8], Paul Cohen proved that the negation of CH is relatively consistent with ZFC. This result, jointly with that proved by Kurt Gödel in 1940 [20]—that GCH is also relatively consistent with ZFC—entails that neither CH nor GCH are provable or refutable from the axioms of ZFC.

Cohen's result came many years after Gödel's incompleteness theorems [19], which imply that there is a sentence in the language of set theory whose truth is not decidable by ZFC. But the enormous surprise was that there are undecidable sentences which are not specifically constructed as a Gödel's sentence; in particular, there is one as simply stated and well known as CH.

There are many mathematical and philosophical issues connected to this outcome. The one which interests us here concerns the consequences it has for ZFC's models: it entails that if ZFC is consistent at all, then it admits a huge variety of different models, where CH and CGH are either true or false and, more generally, the power set class-function (namely $F : \mathfrak{Reg} \to \mathfrak{Reg};\ F(\kappa) = 2^\kappa$, where $\mathfrak{Reg}$ is the class of regular cardinals) behaves in almost arbitrary ways (see below on the results of William Easton). This means that ZFC's axioms leave the von Neumann universe of sets V—which is recursively defined by appealing to the power set operation ($V = \bigcup_\alpha V_\alpha$, with α an ordinal and $V_\alpha = \bigcup_{\beta<\alpha} \mathscr{P}(V_\beta)$)—hugely indeterminate: they are compatible, for example, both with the identification of V with Gödel's constructible universe L (which is what the axiom of constructibility '$V = L$' asserts, by, then, deciding GCH in the positive), and with the admission that in V the values of 2^κ are as large as desired, which makes V hugely greater than L. The question is whether this indetermination of the size of V_α versus the size of L_α can be somehow limited for some sort of cardinals, i.e. for some values of α. The results we mention below show that this is so for singular cardinals, and even, as we said above, that V is asymptotically determined at singular cardinals by its features at the smaller regular cardinals.

To explain this better, we begin with a result by Easton [16], who, shortly after Cohen's result and building on earlier results of Robert Solovay [45], proved that for regular cardinals the indetermination of the values of the power set function is even stronger than the Cohen's result suggests: for any non-decreasing class-function $F : \mathfrak{Reg} \to \mathfrak{Reg}$ defined in an arbitrary model of ZFC so that $\mathrm{cf}(F(\kappa)) > \kappa$ for all κ, there is an extension to another model that preserves both cardinals and cofinalities and in which $2^\kappa = F(\kappa)$, for any regular cardinal κ. This implies that in ZFC no statement about the power set (class)-function[1] on the regular cardinals other than '$\kappa \leq \lambda \implies 2^\kappa \leq 2^\lambda$' and '$\mathrm{cf}\,(\kappa) < \mathrm{cf}\,(2^\kappa)$' can be proved.

It is important to notice that singular cardinals are excluded from Easton's result. Just after the result was obtained, it was felt that this restriction was due to a technical problem which could be overcome in the future. But what became clear later is that this restriction is due to deep differences between regular and singular cardinals. Indeed, many results attesting to this soon followed. In particular, what these results eventually showed is that the power set class-function behaves much better at singular cardinals than it does at regular ones. While the above quoted results by Gödel, Cohen and Easton imply that the value of the power set function can be decided in ZFC for neither regular nor singular cardinals, as not even $2^{\aleph_0}$ has an upper bound there, it turns out that one can do the next-best thing and show in ZFC that the value of 2^κ for any singular κ is conditioned on the values of 2^λ for the regular λ less than κ. This entails that the size of $V_{\kappa+1}$ is, in turn, conditioned by that of that of V_λ for $\lambda \leq \kappa$.

Already by 1965 and 1973 respectively, Lev Bukovský [5] and Stephen H. Hechler [21] had proved, for example, that in ZFC if κ is singular and 2^λ is eventually constant for $\lambda < \kappa$, then 2^κ is equal to this constant. Therefore the value of 2^κ is entirely determined by the values of the power set function below κ. An infinite cardinal λ is said to be *strong limit* if for any $\theta < \lambda$ we have $2^\theta < \lambda$ (in particular, it follows that such a cardinal is limit). Note that strong limit cardinals, and in particular, strong limit singular cardinals, exist in any universe of set theory: an example is given by $\beth_\omega$. Solovay [46] proved that for any κ which is larger or equal to a strongly compact cardinal (a large cardinal λ characterised by having a certain algebraic property that is not essential to explain here, namely that any λ-complete filter can be extended to a λ-complete ultrafilter), we have $2^\kappa = \kappa^+$. In other words, GCH holds above a strongly compact cardinal. This result, of course, is only interesting if there exists a strongly compact cardinal. In fact this result was obtained as part of an investigation started earlier by Dana Scott [36], who investigated the question of what kind of cardinal can be the first cardinal failing GCH, that is, what properties must have a cardinal κ such that $2^\kappa > \kappa^+$, but such that $2^\theta = \theta^+$, for all infinite cardinals $\theta < \kappa$. What Solovay's result shows is that such a cardinal cannot be strongly compact.

---

[1] According to the common abuse of notation, we call F 'power set function', even though it is in fact a class-function.

This result led Solovay to advance a new hypothesis, according to which, for singular cardinals, his own result does not depend on the existence of a strongly compact cardinal. In other words, the hypothesis is that in ZFC, every singular strong limit cardinal κ satisfies $2^\kappa = \kappa^+$. The heart of it is the following implication called the 'Singular Cardinal Hypothesis':

$$2^{\mathrm{cf}(\kappa)} < \kappa \implies \kappa^{\mathrm{cf}(\kappa)} = \kappa^+, \tag{SCH}$$

for any cardinal κ. Indeed, for definition, the antecedent implies that κ is a singular cardinal, so that SCH states that $\kappa^{\mathrm{cf}(\kappa)} = \kappa^+$, for any singular cardinal κ for which this is not already ruled out by $2^{\mathrm{cf}(\kappa)}$ being too big. On the other hand, if κ is a strong limit cardinal, then it follows from the elementary results mentioned in the previous section that $\kappa^{\mathrm{cf}(\kappa)} = 2^\kappa$ (see [27], p. 55), so that the consequent reduces to '$2^\kappa = \kappa^+$'. Hence, SCH implies that the power set operation is entirely determined on the singular strong limit cardinals, since GCH holds for any such cardinal.

In a famous paper appearing in 1975 [44], Jack Silver proved that if κ is a singular cardinal of uncountable cofinality, then κ cannot be the first cardinal to fail GCH. A celebrated and unexpected counterpart of this result was proved by Menachem Magidor shortly afterwards [31]. It asserts that in the presence of some rather large cardinals, it is consistent with ZFC to assume that $\aleph_\omega$ is the first cardinal that fails GCH. This, of course, implies that the condition that κ has uncountable cofinality is a necessary condition for Silver's result to hold. But it also implies that SCH fails and that the power set function at the strong limit singular cardinals does not always behave in the easiest possible way.

Another celebrated theorem proved shortly after the work of Silver is Jensen's Covering Lemma [11], from which it follows that if there are no sufficiently large cardinals in the universe, then SCH holds. To be precise, this lemma implies that SCH holds if $0^\sharp$ does not exist. (It is probably not necessary here to define $0^\sharp$, but let us say that it is a large cardinal whose existence would make V be larger than L, whereas its nonexistence would make V be closely approximated by L.)

Further history of the problem up to the late 1980s is quite complex and involves notions that are out of the scope of ZFC and, a fortiori out of the scope of our paper. Details can be found, for example, in the historical introduction to [43]. Insofar as our interest here is to focus on the results that can be proved in ZFC, we confine ourselves to mention a surprising result proved by Fred Galvin and András Hajnal in 1975 [17]. By moving the emphasis from GCH to the power set function as such, they were the first to identify a bound in ZFC for a value of this function, namely for the value it would take on a strong limit singular cardinal with uncountable cofinality. Let κ be such a cardinal, then what Galvin and Hajnal proved is that $2^\kappa < \aleph_\gamma$, where $\gamma = (2^{|\alpha|})^+$ for that α for which $\kappa = \aleph_\alpha$. As the comparison with the two results of Silver and Magidor mentioned above makes clear, singular cardinals with countable and uncountable cofinality behave quite differently. There were no reasons in principle, then, to think, that Galvin and Hajnal's result would extend to singular cardinals with countable cofinality and the state of the matters stood still for many years.

Fast forward, and we arrive at a crowning moment in our story, namely to the proof, by Saharon Shelah in the late 1980s, of the following unexpected theorem, put forward in [43]:

$$[\forall n \left(n < \omega \implies 2^{\aleph_n} < \aleph_\omega\right)] \implies 2^{\aleph_\omega} < \aleph_{\omega_4}. \tag{1}$$

Shelah's theorem is, in fact, more general than the instance we quoted, which nevertheless perfectly illustrates the point. If $\aleph_\omega$ is a strong limit, then the value of the power set function on it is bounded. In every model of ZFC, Shelah's theorem extends to the countable cofinality the result of Galvin and Hajnal, obtains a bound in terms of just the $\aleph$-function (unlike the Galvin–Hajnal theorem which uses the power set function), and shows that in spite of Magidor's result (which shows that SCH can fails at singular strong limits cardinals of countable cardinality), even at such cardinals a weak form of SCH holds, namely the value of the power set function is bounded.

Shelah's theorem is proved by discovering totally new operations on cardinals, called 'pcf' and 'pp', which are meaningful for singular cardinals and whose values are very difficult to change by forcing. In many instances it is not even known if they are changeable to any significant extent. It would be much too complex for us to describe these operations here but the point made is that even though ZFC axioms are quite indecisive about the power set operation in general, they are quite decisive about it at the singular cardinals and this is because they prove deep combinatorial facts about the operations pcf and pp. The field of research concerned with the operations pcf and pp is called the 'pcf theory'.

3 Some Contemporary Results

The foregoing results have been known to mathematicians for a while but they do not seem to have influenced the literature in philosophy very much. The purpose of this article is to suggest that they have some interest for our philosophical views about ZFC and, more generally, set theory. Before coming to it, however, let us make a short detour in the realm of some more recent results which further illustrate the point. These results, to which this section is devoted, deal with mathematical concepts which are rather advanced; it would distract from the point to present them in a self-contained manner. Those readers who are not at ease with these concepts can safely skip the present section, taking it on trust that contemporary research continues to prove that singular cardinals have quite peculiar features, and that the mathematical universe at such cardinals exhibits much less indetermination than at the regular cardinals. This is the view that we shall discuss in Sect. 4.

Let us begin by observing that the emphasis of the recent research on singular cardinals has moved from cardinal arithmetic to more combinatorial questions. We could say that what recent research on singular cardinals is concerned with is combinatorial SCH: rather than just looking at the value of 2^κ for a certain cardinal κ, one considers the "combinatorics" of κ, namely the interplay of various appropriate properties $\varphi(\kappa)$ of it. An example of such a property might be the existence of a certain object of size κ, such as a graph (see below on graphs) on κ with certain properties, or the existence of a topological or a measure-theoretic object of size κ, in the more complex cases. One may think of κ as a parameter here. Then the relevant instance of combinatorial SCH would say that the property $\varphi(\kappa)$ depends only on the fact that $\varphi(\theta)$ holds at all $\theta < \kappa$. The question can be asked more generally, what about the relevant property of κ can be proved in ZFC, knowing that the property holds all $\theta < \kappa$.

Concerning the former aspect of such a question, that concerned with what can be proved in ZFC, a celebrated singular compactness theorem has been proved by Shelah in [40]. Shelah's book [43] presents, moreover, many applications of pcf theory to deal with this aspect of the question. The latter aspect of the question—namely the forcing counterparts of the former—appeared only later, due to the enormous difficulty of doing even the simplest forcing at a singular cardinal and the necessity (by the Covering Lemma) of using large cardinals, for performing this task. One of the early examples is [14].

To illustrate this sort of research, let us concentrate on one sample combinatorial problem, which has to do with one of the simplest but most useful notions in mathematics, that of a graph.

A graph is a convenient way to represent a binary relation. Namely, a graph (V, E) consists of a set V of vertices and a set $E \subseteq V \times V$ of edges. Both finite and infinite graphs are frequently studied in mathematics and they are also used in everyday life, for example to represent communication networks. Of particular interest in the theory of graphs is the situation when one graph G is subsumed by another one H, in the sense that one can find a copy of G inside of H. This is expressed by saying that there is an embedding from G to H. Mathematically speaking, this is defined as follows.

Definition 1 Suppose that $G = (V_G, E_G)$ and $H = (V_H, E_H)$ are graphs and $f : G \to H$ is a function. We say that f is a *graph homomorphism*, or a *homomorphic embedding* if f preserves the edge relation (so $a\ E_G\ b$ implies $f(a)\ E_H\ f(b)$ for all $a, b \in V_G$) but it is not necessarily 1-1. If f is furthermore 1-1, we say that f is a *weak embedding*. If, in addition, f preserves the non-edge relation (so $a\ E_G\ b$ holds iff $f(a)\ E_H\ f(b)$ holds), we say that f is a *strong embedding*.

Graph homomorphisms are of large interest in the theory of graphs and theoretical computer science (see for example [23] for a recent state-of-the-art book on graph homomorphisms). The decision problem associated to the graph homomorphism, that is, deciding if there is a graph homomorphism from one finite graph into another, is NP-complete (see Chap. 5 of [23], which makes the notion also interesting in computer sciences).

Of particular interest in applications is the existence of a universal graph. If we are given a class $\mathscr{G}$ of graphs, we say that a certain graph G^* is universal for $\mathscr{G}$ if every graph from $\mathscr{G}$ admits a homomorphic embedding into G^*. Of course, variants of this relation can be obtained by replacing homomorphic embedding with weak or strong embedding, as defined in Definition 1. The combinatorial question that we shall survey is that of the existence of universal graphs of a fixed size κ in various contexts.

To begin with, ZFC proves that there is a unique up to isomorphism graph G^* of size $\aleph_0$. This is known as a Rado graph (or, also, random or Erdös–Rényi graph), and it satisfies that for every finite graph G and every vertex v of G, every strong embedding of $G \setminus \{v\}$ into G^* can be extended to a strong embedding of G into G^*. As a consequence, G^* strongly embeds all countable graphs. This graph was discovered independently in several contexts, starting from the work of Ackermann in [2], but its universality properties were proved by Rado in [33].

Under the assumption of GCH, from the existence of saturated and special models in first-order model theory (see [7]), it follows that a universal graph exists at every infinite cardinal κ. In particular, the assumption that $\lambda < \kappa \implies \kappa^\lambda = \kappa$ entails that there is a saturated, and consequently universal, graph of size κ.

When we move away from GCH, the existence of universal graphs becomes a rather difficult problem. Shelah mentioned in [41] a result of his (for the proof see [30] or [12]), namely that adding $\aleph_2$ Cohen reals to a model of CH destroys any hope of having a universal graph of size $\aleph_1$. This does not only mean that there is no universal graph in this model, but also that, by defining *the universality number* of a family $\mathscr{G}$ of graphs as the smallest size of a subfamily $\mathscr{F}$ of $\mathscr{G}$ such that every element of $\mathscr{G}$ embeds into a member of $\mathscr{F}$, we have that in the above model the universality number of the family of graphs of size $\aleph_1$ is the largest possible, namely $2^{\aleph_1}$. More generally, one can state the following theorem:

Theorem 1 (Shelah, see [30] or [12]) *Suppose that $\lambda < \kappa \implies \kappa^\lambda = \kappa$ and let $\mathbb{P}$ be the forcing to add λ many Cohen subsets to κ (with $\mathrm{cf}(\lambda) \geq \kappa^{++}$ and $\lambda \geq 2^{\kappa^+}$). Then the universality number for graphs on κ^+ in the extension by $\mathbb{P}$ is λ.*

Using a standard argument about Easton forcing, we can see that it is equally easy to get negative universality results for graphs at a class of regular cardinals:

Theorem 2 *Suppose that the ground model V satisfies GCH and $\mathscr{C}$ is a class of regular cardinals in V, while F is a non-decreasing function on $\mathscr{C}$ satisfying that for each $\kappa \in \mathscr{C}$ we have $\mathrm{cf}(F(\kappa)) \geq \kappa^{++}$. Let $\mathbb{P}$ be Easton's forcing to add $F(\kappa)$ Cohen subsets to κ for each $\kappa \in \mathscr{C}$. Then for each $\kappa \in \mathscr{C}$ the universality number for graphs on κ^+ in the extension by $\mathbb{P}$ is $F(\kappa)$.*

The proofs of these results are quite easy. In [41], Shelah emphasizes this by claiming that "The consistency of the non-existence of a universal graph of power $\aleph_1$ is trivial, since, it is enough to add $\aleph_2$ generic Cohen reals". He focuses, indeed, on a much more complex proof, that of the consistency of the existence of a universal graph at $\aleph_1$ with the negation of CH. He obtained such a proof in [42], while Mekler obtained a different proof of the same fact in [32]. Insofar as $\aleph_0$ is regular, $\aleph_1$ is the successor of a regular cardinal. Other successors of regular cardinals behave in a similar way, although neither Mekler's nor Shelah's proof seems to carry over from $\aleph_1$ to larger successors of regulars. A quite different proof, applicable to larger successors of regulars but proving a somewhat weaker statement, was obtained by Džamonja and Shelah in [15]: they proved that assuming that it is relatively consistent with ZFC that the universality number of graphs on κ^+ for an arbitrary regular κ is equal to κ^{++} but 2^κ is as large as desired.

All these results only concern regular cardinals and their successors, and leave open the question for singular cardinals and their successors. Positive results analogous to the one just mentioned by Džamonja and Shelah were obtained by Džamonja and Shelah, again, in [14], for the case where κ is a singular cardinal of countable cofinality, and by Cummings, Džamonja, Magidor, Morgan and Shelah in [9], for the case where κ is a singular cardinal of arbitrary cofinality. The most general of their results can be stated as follows:

Theorem 3 (Cummings et al. [9]) *If κ is a supercompact cardinal, $\lambda < \kappa$ is a regular cardinal and Θ is a cardinal with* $\mathrm{cf}(\Theta) \geq \kappa^{++}$ *and $\kappa^{+3} \leq \Theta$, then there is a cardinal preserving forcing extension in which* $\mathrm{cf}(\kappa) = \lambda$, $2^\kappa = 2^{\kappa^+} = \Theta$ *and in which there is a universal family of graphs on κ^+ of size κ^{++}.*

Further recent results of Shelah (private communication) indicate that the universality number in the above model should be exactly κ^{++}. These results concern successors of singular cardinals, which themselves are, of course, regular. The situation for singular cardinals themselves is different; in particular, no forcing notion can operate on them. We do not have any general results about graphs on such cardinals, but here is a result showing that in specific classes of graphs, the existence of a universal element at singulars is simply ruled out by the axioms of ZF (not even the full ZFC is needed):

Theorem 4 (Džamonja [13]) *(ZF) Suppose that κ is a cardinal of cofinality ω. Then, for any $\lambda \geq \kappa$ in ZF, there is no universal element in the class of graphs of size λ that omit a clique of size κ, under graph homomorphisms, or the weak or the strong embeddings.*

This survey of the graph universality problem shows in a specific example the phenomenon of the change in the combinatorial behaviour between the three kinds of cardinals: successors of regulars, successors of singulars and, finally, singular cardinals. At successors of regulars combinatorics is very independent of ZFC, so that simple forcing, without use of large cardinals, allows us to move into universes of set theory which have very distinct behaviours. At the successor of a singular cardinal, we can move from L-like universes only if we use large cardinals (as we know by Jensen's Covering, mentioned above), and this shows up in combinatorics in the necessity to use both large cardinals and forcing to obtain independence results.

This independence is in fact limited (as in the example of Shelah's pcf theorem quoted above). Finally, at singular cardinals, combinatorics tends to be completely determined by ZFC, or even by ZF, as in the example of Theorem 4.

In connection with this theorem, it is interesting to note that in the absence of the Axiom of Choice, it is possible that every uncountable cardinal is singular of countable cofinality. To be exact, Gitik proved in [18] that from the consistency of ZFC and arbitrarily large strongly compact cardinals, it is possible to construct a model of ZF in which all cardinals have countable cofinality. Therefore, if one is happy to work with ZF only, then one has the choice to move to a model in which only singular cardinals exist and they only have countable cofinality. In such a model, combinatorics becomes easy and determined by the axioms, at least in the context of the questions that have been studied, such as the graph universality problem.

4 Philosophical Remaks

Mathematical platonism is often presented as the thesis that mathematical objects exist independently of any sort of human (cognitive, and/or epistemic) activity, and it is taken to work harmoniously with a realistic semantic view, according to which all we can say in mathematics (i.e. by using a mathematical language) is either true or false, to the effect that all that has been (unquestionably) proved is true, but not all that is true has been (unquestionably) proved or can be proved (because of various forms of incompleteness of most mathematical theories).

Both claims are, however, quite difficult to support and are, in fact, very often supported only by the convenience of their consequences, or, better, by the convenient simplicity of the account of mathematics they suggest, and because they provide a simple explanation of the feeling most mathematicians (possibly all) have that something external to them resists their intuitions, ideas, programs, and conjectures, to the effect that all that they can frame by their thoughts or their imagination must have, as it were, an external, independent approval, before having its place among mathematical achievements. Hence, an interesting philosophical question is whether there can be weaker claims that have similarly convenient consequences and that can be more easily positively supported, either by evidence coming from mathematical practice, or by more satisfactory metaphysical assumptions, or, better, by both.

It is our opinion that such claims can reasonably be formulated. In short, they are the following: (*i*) there are ways for us to have epistemic *de re* access to mathematical objects; (*ii*) we are able to prove truths about them, though others are still not proved or are unprovable within our most convenient theories (which are supposed to deal with these objects). Claim (*i*) means that there are ways for us to fix intellectual contents which are suitably conceived as individuals that mathematics is dealing with, in such a way that we can afterwards (that is, after having fixed them) ascribe properties and relations to these individuals. Claim (*ii*) means that some of our ascriptions of property and relations to these individuals result in truths, in the sense that they somehow comply with the content we have afterwards fixed, and, among them, some can be, and in many cases have been, provably established, though others are still not so or cannot be so within the relevant theories.

The phrase '*de re*' in claim (*i*) belongs to the philosophical lexicon. It is currently used in opposition to '*de dicto*' to point out a distinction concerning propositional attitudes, typically belief (or knowledge). Believing that something is P can mean believing either that there is at least one thing that is P or that some specific thing is P. In the former case the belief is *de dicto*; in the latter *de re*. If the relevant thing is t, a suitable way to unambiguously describe the second belief is saying that of t, it is believed that it is P. This makes clear that the subject of a *de re* propositional attitude is to be identified independently from ascribing to it what the relevant proposition ascribes to it. Hence, its being P cannot be part of what makes it what it is. This is not enough, however, since for the attitude to be *de re*, the identification has to be stable under its possible variations. If Mirna believes that t is the only thing that is Q, her believing of t that it is P is the same as her believing it of the Q. But Marco can believe that the only thing that is Q is s (distinct from t). So his believing of the Q that it is P is quite distinct from Mirna's belief that it is so. Hence neither beliefs are *de re*. This makes clear that the identification of the subject of a *de re* attitude is to be independent of the attitude itself or, even, of any sort of attitude (since different attitudes can compose each other's). This is why the most straightforward examples of *de re* attitudes concern empirical objects ostensively, or pre-conceptually identified in one way or another.

This has not prevented philosophers from appealing to the *de re* versus *de dicto* distinction in relation to mathematics. In particular, a rich discussion has concerned the possibility of using appropriate sorts of numerals for directly referring to natural numbers while having a *de re* attitude towards them. Diana Ackerman has considered that "the existence of [natural] numbers is a necessary condition for anyone's having *de re* propositional attitudes toward them" ([1], p. 145). Granted their existence, Tyler Burge has wondered whether we can have "a striking relation to [...][a natural number] that goes beyond merely conceiving of it or forming a concept that represents it", and answered that this is so for small such numbers, since "the capacity to represent [...][them] is associated with a perceptual capacity for immediate perceptual application in counting" ([6], pp. 70–71). Saul Kripke has gone far beyond this, by suggesting a way to conceive natural numbers that makes decimal numerals apt to "reveal[...] their structure" ([39], p. 164; [29]). For him, natural numbers smaller than 10 are the classes of all n-uples ($n = 0, 1, \dots, 9$), while those greater than 9 are nothing but finite sequences of those smaller than 10. This makes decimal numerals, or, at least, short enough ones, work as "buckstoppers" (i.e. they are such that it would be nonsensical asking which number is that denoted by one of them, in opposition to terms like 'the smallest perfect number', denoting the natural number whose buckstopper is 'six'), and so allow direct reference to them. By dismissing such a compositional conception of natural numbers, Jan Heylen [25] and Stewart Shapiro [38] have respectively submitted that Peano numerals (the numerals of the form '$0^{\prime\cdots\prime}$', written using only the primitive symbols for zero and the successor relation in the language of Peano Arithmetic) and unary numerals (mere sequence of n strokes used to denote the positive natural number n) provide canonical notations allowing *de re* knowledge of natural numbers. Finally, Jody Azzouni [3] has argued that the existence of natural numbers is not required for having "*de re* thought" about them, since such a thought can be "empty".

Our use of '*de re*' in claim (*i*) differs from all these uses in that the *de re* versus *de dicto* distinction has a much more fundamental application in our account of mathematics. Far from merely concerning our way of denoting natural numbers so to identify them in such a way to make *de re* propositional attitudes towards them possible, granted their existence, or our *de re* thought about them empty, granted their nonexistence, it concerns our way of fixing mathematical objects so as to confer existence to them. In our view these objects are, indeed, nothing but contents of (intentional) though, whose existence just depends on the way they are fixed. Here is how we see the matter.

There are many ways of fixing intellectual contents, which, in appropriate contexts, are (or can be) suitably conceived as individuals. A liberal jargon can refer to these contents as abstract objects. If this jargon is adopted, the claim that mathematics deals with abstract objects becomes quite trivial, and can neither be taken as distinctive of a platonist attitude, nor can provide any characterisation of mathematics among other intellectual enterprises. In a much more restrictive jargon, for something (i.e. the putative reference of a term or description) to count as an object, it has to exist. Under this jargon, the claim that mathematics deals with abstract objects becomes much more demanding, overall if it is either required that these objects are self-standing or mind-independent, or if it is supposed that nothing can acquire existence because of any sort of intellectual (intentional) act. The problem, then, with this claim is that it becomes quite difficult to understand what 'to exist' can mean if referred to abstract contents. What we suggest is reserving the term 'abstract object' to intellectual contents suitably conceived as individuals and so fixed, in an appropriate context, so as to admit *de re* epistemic access, this being conceived, in turn, as the apprehension of them making *de re* attitudes towards them possible. We submit that, once this is granted, the claim that mathematics deals with abstract objects becomes both strong enough and distinctive, so as to provide the ground for an appropriate account of mathematics.

Mathematics traditionally admits different modalities for fixing intellectual contents. The French philosopher Jean-Michel Salanskis [34, 35] suggested to distinguish two basic ways of doing it: constructively and correlatively.

The former way has a more limited application, but can be taken, in a sense, as more fundamental. Peano's numerals can, for instance, be quite simply fixed constructively by stating that: (*i*) the sign '0' is a Peano's numeral; (*ii*) if the sign 'σ' is such a numeral, then the sign 'σ'' is such a numeral, too; (*iii*) nothing else is such a numeral. Similarly, unary numerals can be constructively fixed by stating that: (*i*) the sign '|' is such a unary numeral; (*ii*) if the sign 'σ' is such a numeral, then the sign 'σ |' is such a numeral, too; (*iii*) nothing else is such a numeral. These are numerals, not numbers, however. And is clearly unsuitable to use the same pattern to define natural numbers. Suppose it were stated that: (*i*) 0 is a natural number; (*ii*) if σ is such a number, then σ' is such a number; (*iii*) nothing else is such a number. It would have not been established yet that there is no natural number n such that $0 = n'$, or

$n = n'$. To warrant that this is so, it would still be necessary to impose appropriate conditions to the successor function $-'$, which cannot be done constructively. To overcome the difficulty, one could have recourse to a trick: stating the natural numbers are the items that Peano numerals denote, or positive such numbers the items that unary numerals denote, in such a way that distinct such numerals denote distinct such numbers. This would make Peano's numerals directly display the structure of natural numbers, and unary ones that of positive natural numbers, so providing a canonical notation for these numbers allowing direct reference to them, in agreement to Heylen's and Shapiro's proposals. But this would be dependent on the informal notion of denotation. Supposing that we have the necessary resources for handling this notion without ambiguity, this would allow us to fix natural numbers almost constructively. Once this is done, one could look at these numbers as such, and try to disclose properties they have and relations they bear to each other's. Making it in agreement with mathematical requirements of rigor asks both for further definitions and the fixation of inferential constraints or rules, typically of an appropriate codified, if not formal, language. What is relevant for illustrating our point, is, however, not this, but rather that that we can do both things in such a way to keep the reference steady to the contents previously fixed as just said: it is on them that we define the relevant properties and relations; and it is to speak of them that we establish the appropriate inferential constraints, and fashion (or adopt) the appropriate language, which allows us to say of them, or some of them, that they are so and so. This should give a rough idea of the intellectual phenomenon we want to focus on by speaking of *de re* epistemic access.

More importantly, we could observe that once appropriate intellectual contents are fixed constructively, one can also try to capture them correlatively, that is, through an axiomatic implicit definition. This can be done somehow informally, or by immersing the definition within a formal system affording both the appropriate language and the appropriate inference rules (or, possibly, allowing to state these rules). In the case of natural numbers, we can, for instance, define them, through Peano axioms, within an appropriate system of predicate logic, and we could conceive of doing that with the purpose of characterizing correlatively the same contents previously fixed constructively, so as that each of them provide the reference for a singular term appropriately introduced within the adopted language, and that they provide, when taken all together, the domain of variation and quantification of the individual variables involved in the definition.

The predicate system adopted can be both first- or higher-, typically second-, order. There is, however, a well-known difference among the two cases: while Peano second-order arithmetic (or PA2, for short) is categoric (with respect to the subjacent set theory), by a modern reformulation of Dedekind's argument [10], Peano first-order arithmetic (or PA1, for short) is not, by an immediate consequence of the Löwenheim-Skolem's theorem [7]. This suggests that the verb 'to capture' is not to be understood in the same way in both cases. In the second-order case, it means that the relevant axioms determine a single structure (up to isomorphism), whose elements are intended to be the natural numbers, identified with the same objects previously fixed constructively. In the first-order case, it means that these axioms describe a class of non-isomorphic structures, all of which include individuals that behave, with respect to each other's, in the same way as the elements of this structure do, and that we can then intend, again, as the same objects previously fixed constructively.

Both in the usual platonist tongue, and in our amended one, we could say that the limited expressive power of a first-order language makes it impossible to univocally describe the natural numbers by means of such a language: to do it, a second-order language is needed (and it suffices). Still, the verb 'to describe' should be understood differently in the two cases: while in the former case it implies that that these numbers are self-standing objects that are there as such, independently of any intellectual achievement, in the latter case, it merely implies that these objects have been previously fixed. Hence, if no previous definition were admitted or considered, the verb 'to fix' should be used instead. What should, then, be said is that the limited expressive power of a first-order language makes it impossible to univocally fix the natural numbers by means of such a language. (Of course, the relativisation of the categoricity of PA2 to a given model of set-theory makes the usual platonist tongue appropriate only insofar as it is admitted that this model reflects the reality of the world of mathematical objects, which, in presence of the strong non-categoricity of ZFC requires a further act of faith. But on this, later.)

The difference between and first- and the second-order case is not limited to this, however. Another relevant fact is that the language of PA1 is forced to include, together with the primitive constants used to designate the number zero and the successor relation, also two other primitive constants used to designate addition and multiplication. (Though versions of PA1 often adopt a language including a further primitive constant used to designate the order relation, this can be easily defined in terms of addition by, then, reducing the number of axioms, albeit increasing the syntactical complexity of some proofs.) The only primitive constants which are required to be included in the language of PA2 are, instead, those used to designate the number zero and the successor relation: addition and multiplication (as well as order), can be recursively defined in terms of zero and successor. Hence, whereas Peano second-order axioms (implicitly) define a structure $\langle \mathbf{N}, ' \rangle$ Peano first-order axioms define uncountably many distinct structures $\langle \mathbf{N}, ', +, \times \rangle$. It remains the fact, nevertheless, that the former structure is reflected within any one of the latter ones. Hence, if we admit that the axioms of PA2 capture or fix a domain of objects in an appropriate way, there is room to say that PA1 is studying these same objects by weaker logical means, by identifying them as the common elements of uncountably many possible structures $\langle \mathbf{N}, ', +, \times \rangle$, though being unable to provide an univocal characterisation of them.

This should clarify a little better what having epistemic *de re* access to mathematical objects could mean: one could argue that, once natural numbers are captured or fixed by the axioms of PA2 as the elements of $\langle \mathbf{N}, ' \rangle$, one can, again, look at them as such and try to disclose their properties and relations, so as to recover the same property or relation already ascribed to them, and possibly more. This can be done in different ways. By staying within PA2, one can, for example, besides proving the relevant theorems statable in its primitive language, also enrich this language bymeans of appropriate explicit definitions, so as to introduce appropriate constants—as those designating addition multiplication and order—to be used in the relevant proofs. By leaving this theory, one can also try to describe them by using a weaker language, such as a first-order one, and be, then, forced to implicitly define addition and multiplication in them by appropriate axioms, though being unable to reach an univocal description. Other ways for studying these numbers are, of course, at hand. But, for our present purpose, we can confine ourselves to observe that in this latter case (as in many other ones), what we are doing may be appropriately accounted for by saying that, of these very numbers, we claim (by using the relevant first-order language) that they are so and so, or, better, that they form a structure $\langle \mathbf{N}, \,', +, \times \rangle$.

There is a quite natural objection one could address to these views. One could remember that, as any other second-order theory, PA2 is syntactically incomplete, to the effect that some statements that are either true or false in its unique model are neither provable nor disprovable in it, and there is, then, no way (or at least no mathematically appropriate way) for us to know whether they are true or false. Hence, one could argue, whatever a *de re* access to natural numbers, as defined by PA2, might be, it cannot be, properly speaking, an epistemic access, since there are not only things about these numbers that we do not know, but also things that we cannot know. We think this objection misplaced, since something analogous also occurs for genuine empirical objects. Take the chair you sit on (if any): there are many properties that we suppose (at least from a realist perspective) that it does or does not have, about which even our best theories and the information we are in place to obtain are insufficient to make a decision. This should not imply, it seems to us, that you have no knowledge of that chair. Of course, we could always change our theories or improve them if we considered that deciding some questions that we know to be undecidable within them is relevant. In the same way, if we were considering (or discovering) that there are some relevant statements about natural numbers which are provably undecidable in PA2, we could try to add axioms to the effect of provably deciding these statements. But allowing this possibility does not imply that we do not have *de re* epistemic access to these numbers as fixed by PA2, while working on them either within or outside it. All that is required for it is that there is a suitable sense in which we can say that on these numbers (as independently fixed) we can define some properties or relations within this theory, or of these numbers we can claim this or that outside the theory.

Something similar to what happens with PA2 also happens with Frege arithmetic (or FA, for short), namely full (dyadic) second-order logic plus Hume's Principle (see Wright [49] or [4], especially Sect. II). The role played by natural numbers in the former case is played by the cardinal ones (understood as numbers of concepts) in the latter case. Once a particular cardinal number, typically the number of an (or the) empty concept is identified with 0, and an appropriate functional and injective relation is defined on these numbers so as to play the role of the successor relation, one can select the natural numbers among the cardinal ones, as being 0 together with all its successors. One can then capture or fix the natural numbers without appealing to addition and multiplication on them (and no more on order, at least explicitly).But now there is even more: these numbers can be captured or fixed by selecting them among items which are fixed, in turn, by appealing neither to a designated item like 0, nor to a certain dyadic relation, like the successor relation. Of the cardinal numbers, one could, then, say, that some of them are the natural ones and can be studied as such with other appropriate means.

It is easy to see that, as opposed to PA2, FA is not categoric (with respect to the subjacent set theory). This merely depends on the presence in some of its models of objects other than cardinal numbers, which can be absent from others. Still, FA interprets PA2 (this is generally known as Frege's theorem: see [22], for example), and a result of relative categoricity can also be proved for FA ([47], prop. 14; [48], pp. 573–574): any two models of it restricted to the range of the number-of operator are isomorphic (with respect to the subjacent set theory). This might make one think that a form of categoricity (with respect to the subjacent set theory) is essential for allowing *de re* epistemic access to mathematical objects, i.e. that the only intellectual contents suitably conceived as mathematical objects that we can take to have *de re* epistemic access to are those fixed within a theory endowed with an appropriate form of categoricity (with respect to the subjacent set theory).

This is not what we want to argue for, however. The previous example of the constructive definition of positive natural numbers should already make it clear. Another, quite simple example is the following: when we define the property of being a prime number within PA1, we do it on the natural numbers in such a way that we can say that on these numbers we define this property; if the definition is omitted, many usual theorems of PA1 can no longer be proved, of course, but this does not change anything to many other theorems still concerned with natural numbers as defined within this theory. These two examples are different from each other, and both different from that given by the access to natural numbers as defined within PA2. That provided by the definition of prime numbers within PA1 is only an example of *de re* epistemic access internal to a given theory, which reduces, in fact, to nothing more than the possibility of performing an explicit definition within this very theory. Claiming that we have *de re* epistemic access to natural numbers as defined constructively, or to these very numbers as defined correlatively within PA2, when we try to study them in a different context, is quite a different story. Still, there is something similar in the three cases, and this is just what we are interested in underlining here: it is a sort of (relative) stability of intellectual contents counting as mathematical objects, a stability that is made possible by the way these contents are fixed. We do not want to venture here in the (possibly hopeless) tentative of classification of forms of *de re* epistemic access. Still, it seems clear to us that the phenomenon admits differences: both the stability depending on a constructive, or, more generally, informal definition, and that depending on a categorical implicit formal definition are extra-theoretic; the former is strictly intentional, as it were, the latter semantic; that depending on explicit definitions within non-categoric theories is merely syntactic (and, then, intra-theroretic) or restricted, at least, to an informally identified intended model. But the notion of independent existence of mathematical objects, which usual platonism is concerned with, is imprecise enough to make it possible to hope that all these different sorts of stability can provide an appropriate (metaphysically weak) replacement of it in many cases in which platonists use it in their accounts of mathematics.

# # #

But, let it be as it may. The question here is different: what does all this have to do with ZFC, and the results mentioned in Sects. 2 and 3, above?

On the one side, it is clear not only that the categoricity of PA2 and FA is relative to the (inevitably arbitrary) choice of a model of set-theory, and, then, typically, of ZFC, but also that what has been said about PA1, PA2 and FA has a chance to be clear only if set-theory provides us with a clarification of the relevant crucial notions. This is, however, not enough for concluding that whatever philosophical position we could take on natural numbers, and other mathematical objects along the lines suggested above, is necessarily dependent on a preventive account of ZFC. On the one side, we do not need all the expressive and deductive power of ZFC, and a fortiori of whatsoever acceptable extension of it, to make the relevant notions clear. On the other side, it is exactly the high un-categoricity of ZFC that invites us to reason with respect to finite numbers under the supposition that a model of the subjacent set-theory has been chosen, or, even, independently of the preventive assumption that these numbers are sets.

This suggests taking ZFC as an independent mathematical theory—one, by the way, powerful enough to be used (among other things) for studying from the outside the structures formed by the natural numbers, as well as by other mathematical objects, as objects we have a *de re* epistemic access to independently of (the whole of) it. One could then ask whether some sort of *de re* epistemic access to pure sets (conceived as *sui generis* objects implicitly defined by ZFC) is possible or conceivable. The high un-categoricity of ZFC seems to suggest a negative answer. Because it looks like neither this theory as such, nor any suitable extension of it (with the only exception, possibly, of ZFC + '$V = L$', if this might be taken to be a suitable theory, at all) can provide a way to fix pure sets in any appropriate way for allowing *de re* (semantic) epistemic access to them. Upon further reflection, the case can appear, however, not to be so desperate as it seems to be at first glance, and the results mentioned above help us in seeing why this is so.

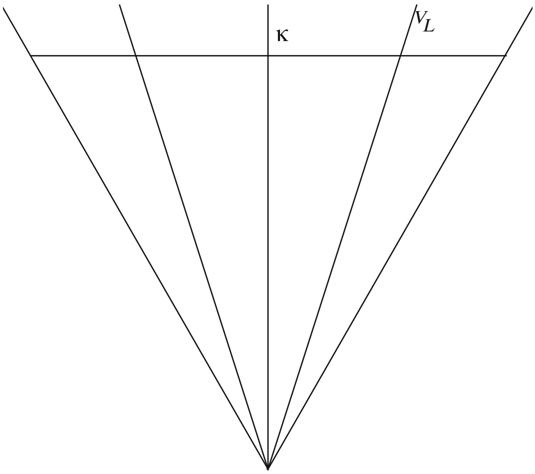
To begin with, one might wonder whether, in analogy to what we have said concerning PA1 and PA2, ZFC could not be taken as studying pure sets as the objects previously fixed in a quasi-categorical way by ZF2, just like PA1 might be taken to do with the natural numbers as (captured or) fixed by PA2.

The problem with this suggestion is that the relations between ZFC and ZF2 are not as illuminating as those between PA1 and PA2. For example, if we fix a level of the cumulative hierarchy of sets, say V_α, then the second-order theory of V_α is simply the first-order theory of $\mathscr{P}(V_\alpha) = V_{\alpha+1}$, hence passing to the second-order does not seem like it has achieved much. However, it is true that formulating ZF in the full second-order logic so as getting ZF2, one achieves what is known as quasi-categoricity. The proof is basically contained in Zermelo [50]. We can describe the situation in more detail although informally, as follows.

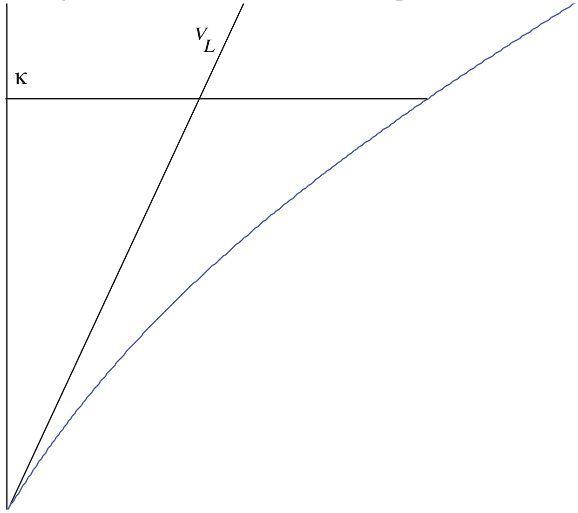
What Zermelo proved for ZF2 is that for any strongly inaccessible cardinal υ which is supposed to exist, there is a single model (up to isomorphism) of ZF2 provided by the structure $\langle V_\upsilon, \in \rangle$. It follows that all theories ZF2 + 'there are exactly n strongly inaccessible cardinal' ($n = 0, 1, 2, \ldots$), or ZF2$_n$, for short, are fully categorical, giving that ZF2 has, modulo isomorphism, as many (distinct) models as there are strongly inaccessible cardinals (recall that V_υ can only include strongly inaccessible cardinals smaller than υ). Of course, in any of these models any statement of the language of ZF2 is either true or false (according to the Tarski's semantic). But, because of the proof-theoretical incompletess of the second-order logic, and, then, of any second-order theory, it is not necessarily decidable. As noted below, this is so also for PA2. The difference is that in these extensions of ZF2, the undecidable statements include some with a clear and unanimously perfectly recognized mathematical significance, namely CH and GCH.

Now, while the problem of deciding GCH (for cardinals greater than $2^{\aleph_0}$) can be seen as intrinsically internal to set theory (both to ZFC and ZF2), this is not so for CG. For, if we admit that there are (necessarily not-constructive) ways to fix real numbers, so as to allow us to have *de re* epistemic access to them (for example within PA2, as originally suggested by Hilbert and Bernays ([24], supplement IV), the problem of deciding CH can be seen as the question of answering the very natural question of how many are such numbers, a question which should, then, be seen as having a definite answer outside set theory (both ZFC and ZF2). The difference is, then, relevant, also from the point of view we are delineating.

Usually, a model V_M of ZFC is diagrammatically represented this way:



where V_L is the model of ZFC + '$V = L$', and the external triangle can coincide with the internal one (which happens if '$V = L$' is true in the model), but not go up to become internal. However, insofar as nothing requires that a model of ZFC have a uniform hierarchic shape, and no significant feature of it is represented by the symmetry of the diagram, we submit that a better representation is the following



where all that is required of the external curve, call it '$\mathscr{C}$', for short, is that it is everywhere increasing (with respect to the line of cardinals, taken as axe) and external or coincident to the internal half straight-line. If this picture is adopted, a model of ZF2 could be depicted in the same way, with the specification that the external curve is univocally determined by the choice of a strongly inaccessible cardinal υ, or by the supposition that there are exactly n such cardinals, which leads to our calling it '$\mathscr{C}_\upsilon$' or '$\mathscr{C}_n$'.

One could, then, advance that (the axioms of) ZF2 plus the choice of a strongly inaccessible cardinal, or (those of) $\mathrm{ZF2}_n$ allow to univocally fix a domain of *sui generis* objects—call it 'the υ-sets' or 'the n-sets'—and that ZFC is studying these very objects with weaker logical means as elements of uncountably many possible structures, being unable to provide an univocal characterisation of them.

This suggests that ZF2, plus the choice of a strongly inaccessible cardinal, or $\mathrm{ZF2}_n$ provide domains of objects we can have a *de re* access to, in the same way as this happens for PA2, that is, not only internally, and so providing a sort of syntactic stability, but also externally, so as to provide a sort of semantic stability: one could argue that, once pure sets are fixed by the relevant (second-order) axioms, one can look at them as such and try to tell (both using a first- or a second-order language) the properties they have or the relations they bear to each other's. Of them, we claim that they form a structure that ZF(C) and all its usual (first-order) extensions try to describe, though being unable to univocally identify.

Still, the relativisation to the choice of a strongly inaccessible cardinal or the admission of the supplementary axiom 'there are exactly n strongly inaccessible cardinals' make the situation much less satisfactory than the one concerned with Peano (first- and second-order) arithmetic: taken as such, ZF2 is not only proof-theoretically incomplete; it is also unable to univocally fix the relevant objects.

This relativisation or admission do not prevent us from ascribing, however, to ZF2 a form of categoricity, since from Zermelo's result "it also follows that every set-theoretical question involving only sets of accessible rank is answerable in ZF2", and, then, in particular, that "all propositions of set theory about sets of reals which are independent of ZFC", among which there is CH, are either true or false in any of its model, though no proof could allow us to establish whether the former or the latter obtains ([26], p. 790). This might be taken as very good news. But a strong objection is possible: it is possible to argue that the truth or falsity of CH in any model of ZF2 does not depend on the very axioms of this theory, but on the consequence relation which is determined by the use of second-order logic and the standard (or full) interpretation of it, or, in other terms, that what makes CH true or false there is not what the axioms of ZF2 genuinely say about sets, but their using second-order variables, semantically interpreted as sets of n-tuples on the fist-order domain. Clearly, this would make second-order logic so interpreted "inadequate for axiomatizing set theory" (see [26], pp. 782 and 790–793, for details).

We do not want enter such a delicate question here. We merely observe that the mathematical results we have expounded above show that there is no need to go second-order to get a limited form of quasi-categoricity. Since these results suggest that ZFC has already (and alone, that is, without any need to appeal to any supplementary axiom) the resources for fixing some of its objects in a better way than it is usually thought. Namely, if we are happy to work at a singular cardinal then much of the combinatorics is determined by what happens at the regular cardinals below, even to the point of fixing the cardinal arithmetic (see Shelah's Theorem 1 quoted above). In some cases, we do not even need to know what happens at the regular cardinals below (see Theorem 4). And if we are happy to be in a world with no Axiom of Choice, we can even imagine that all cardinals are singular, as in the Gitik's model and hence much of the cardinal combinatorics is completely determined by ZF.

Let us look back to the second of the previous figures and suppose that κ is a singular cardinal. What these results suggest is this: if the values of the ordinates of $\mathscr{C}$ are fixed for all regular cardinals λ smaller than κ, i.e. if a single model of ZFC is chosen relatively to all these regular cardinals, then the value of the ordinate of $\mathscr{C}$ for κ is strongly constrained, in the sense that this value can only belong to a determined set (a set, not a class) of values. In other terms, things seem to happen as if the shape of a model of ZFC for the regular smaller than κ strongly conditions the shape of the possible models at κ.

These results could be understood as saying that the non-categoricity of ZFC is, in fact, not as strong as it appears. Even within first-order, the behavior of the universe of sets is fixed enough at singular cardinals to give us some sort of external and semantic *de re* epistemic access to them and their power sets. In particular, once we have given to us all sets of size $< \kappa$ and all their power sets, our choices for κ are quite limited. This offers an image of the universe of sets in which a strong lack of uniuvocality only concerns successor cardinals or uncountable regular limit cardinals, if any (remember that the existence of uncountable regular limit cardinals is unprovable in ZFC). One could say that, at singular limits, ZFC already exhibits a form of categoricity, or, better, that it does it asymptotically, since the ratio of singular cardinals over all cardinals tends to 1 as the cardinals grow. And at the price of working only in ZF we can even imagine to be in the model of Gitik, in which every uncountable cardinal is a singular limit.

Under a realist semantic perspective, according to which all we could say about the universe of sets is either true or false, one could say that this shows that, though ZFC is unable to prove the full truth about this universe, it provably provides an asymptotic description where the singular cardinals are the limits of the asymptotes. This also suggests, however, an alternative and more sober picture, which is what we submit: though there is no sensible way to say what is true or false about the universe of sets, unless truth and falsity are merely conceived as provable truth and falsity, ZFC provides an asymptotically univocal image of the universe of sets around the singular cardinals: the image of a universe to which we can have an external semantic *de re* epistemic access.

Acknowledgements The first author gratefully acknowledges the help of EPSRC through the grant EP/I00498, Leverhulme Trust through research Fellowship 2014–2015 and l'Institut d'Histoire et de Philosophie des Sciences et des Techniques, Université Paris 1, where she is an Associate Member. The second acknowledges the support of ANR through the project *ObMathRe*. The authors are grateful to Walter Carnielli for his instructive comments on a preliminary version of the manuscript, and to Marianna Antonutti-Marfori, Drew Moshier and Rachael Schiel for valuable suggestions.

References

1. Ackerman, Diana. 1978. De re propositional attitudes toward integers. *The Southwestern Journal of Philosophy* 9 (2): 145–153.
2. Ackermann, Wilhelm. 1937. Die Widerspruchsfreiheit der allgemeinen Mengenlehre. *Mathematische Annalen* 114 (1): 305–315.
3. Azzouni, Jody. 2009. Empty *de re* attitudes about numbers. *Philosophia Mathematica*, ser. III, 17 (2):163–188.
4. Boolos, George. 1999. *Logic, Logic and Logic*. Cambridge (Mass): Harvard University Press.
5. Bukovský, Lev. 1965. The continuum problem and powers of alephs. *Commentationes Mathematicae Universitatis Carolinae* 6: 181–197.
6. Burge, Tyler. 2007. Postscipt to "Belief *De Re*". In *Foundations of Mind. Philosophical Essays*, ed. by Tyler Burge, vol. 2, pp. 65–81. Oxford: Clarendon Press.
7. Chung Chang, Chen, and H. Jeremy Keisler. 1990. *Model Theory*, 3rd edn. Amsterdam: North Holland.
8. Cohen, Paul. 1963. The independence of the continuum hypothesis. *Proceedings of the National Academy of Sciences of the United States of America* 50 (6): 1143–1148.
9. Cummings, James, Mirna Džamonja, Menachem Magidor, Charles Morgan, and Saharon Shelah. 2017. A framework for forcing constructions at successors of singular cardinals. TAMS 369 (10): 7405–7441.
10. Dedekind, Richard. 1888. Was sind und was sollen die Zahlen? F. Viewing and Sohn in Braunschweig.
11. Devlin, Keith I., and R.B. Jensen. 1975. Marginalia to a theorem of Silver. In $\models$*ISILC Logic Conference (Proc. Internat. Summer Inst. and Logic Colloq., Kiel, 1974)*, pp. 115–142. Lecture Notes in Math., Vol. 499. Berlin: Springer.
12. Džamonja, Mirna. 2015. The singular world of singular cardinals. In *Logic Without Borders: Essays on Set Theory, Model Theory, Philosophical Logic and Philosophy of Mathematics*,

ed. by Åsa Hirvonen, Juha Kontinen, Roman Kossak, and Andrés Villaveces, pp. 139–146, Boston. De Gryter.

13. Džamonja, Mirna. 2016. Universal infinite clique-omitting graphs. *Sarajevo Journal of Mathematics*, 12 (25): 151–154.
14. Džamonja, Mirna, and Saharon Shelah. 2003. Universal graphs at the successor of a singular cardinal. *Journal of Symbolic Logic* 68: 366–387.
15. Džamonja, Mirna, and Saharon Shelah. 2004. On the existence of universal models. *Archive for Mathematical Logic* 43 (7): 901–936.
16. Easton, William B. 1970. Powers of regular cardinals. *Annals of Mathematical Logic* 1: 139–178.
17. Galvin, Fred, and András Hajnal. 1975. Inequalities for cardinal powers. *Annals of Mathematics* 2 (101): 491–498.
18. Gitik, Moti. 1980. All uncountable cardinals can be singular. *Israel Journal of Mathematics* 35 (1–2): 61–88.
19. Gödel, Kurt. 1931. Über formal unentscheidbare Säztze der Principia Mathematica und verwandter Systeme, I. *Monatshefte für Mathematik und Physik* 38: 173–198.
20. Gödel, Kurt. 1997. *The Consistency of the Continuum-Hypothesis*. Princeton: Princeton University Press.
21. Hechler, Stephen H. 1973. Powers of singular cardinals and a strong form of the negation of the generalized continuum hypothesis. *Z. Math. Logik Grundlagen Math*. 19: 83–84.
22. Heck, Richard G. 2011. *Frege's Theorem*. Oxford: Clarendon Press.
23. Hell, Pavol, and Jaroslav Nešetřil. 2004. *Graphs and homomorphisms*, vol. 28., Oxford Lecture Series in Mathematics and its Applications Oxford: Oxford University Press.
24. Hilbert, David, and Paul Bernays. 1934–1939. *Grundlagen der Mathematik*, vol. 2. Berlin, Heidelberg, New York: Springer. Second edition: 1968–1970.
25. Heylen, Jan. 2014. The epistemic significance of numerals *Synthese*, Published online on September 6th. https://doi.org/10.1007/s11229-014-0542-y.
26. Jané, Ignacio. Higher-order logic reconsidered. In [37], pp. 781–808.
27. Jech, Thomas. 2003. *Set Theory*, 3rd millenium edn. Berlin Heidelberg: Springer.
28. Kojman, Menachem. 2011. Singular Cardinals: from Hausdorff's gaps to Shelah's pcf theory. In *Sets and Extensions in the Twentieth Century* ed. by Dov M. Gabbay, Akihiro Kanamori, and John Woods, vol. 6 of *Handbook of the History of Logic*, pp. 509–558. Elsevier.
29. Kripke, Saul. 1992. *Whitehead Lectures*. Umpupished transcription of Kripke's lectures delivered at Harvard University on May 5th-6th.
30. Kojman, Menachem, and Saharon Shelah. 1992. Nonexistence of universal orders in many cardinals. *Journal of Symbolic Logic* 57 (3): 875–891.
31. Magidor, Menachem. 1977. On the singular cardinals problem. I. *Israel Journal of Mathematics* 28 (1–2): 1–31.
32. Mekler, A. 1990. Universal structures in power $\aleph_1$. *Journal of Symbolic Logic* 55 (2): 466–477.
33. Rado, Richard. 1964. Universal graphs and universal functions. *Acta Arithmetica* 9: 331–340.
34. Salanskis, Jean-Michel. 1991. *L'heméneutique formelle*. Paris: Éditions du CNRS.
35. Salanskis, Jean-Michel. 2008. *Philosphie des mathématiques*. Paris: Vrin.
36. Scott, Dana. 1961. Measurable cardinals and constructible sets. *Bull. Acad. Polon. Sci. Sér. Sci. Math. Astronom. Phys*. 9: 521–524.
37. Shapiro, Stewart. 2005. *The Oxford Handbook of Philosophy of Mathematics and Logic*. Oxford, New York: Oxford University Press.
38. Shapiro, Stewart. 2017. Computing with numbers and other non-syntactic things: *De re* knowledge of abstract. *Philosophia Mathematica*, ser. III, 25 (2): 268–281.
39. Steiner, Mark. 2011. Kripke on logicism, Wittgenstein, and de re beliefs about numbers. In *Saul Kripke*, ed. Alan Berger, 160–176. New York: Cambridge University Press.
40. Shelah, Saharon. 1975. A compactness theorem for singular cardinals, free algebras, Whitehead problem and transversals. *Israel Journal of Mathematics* 21 (4): 319–349.
41. Shelah, Saharon. 1984. On universal graphs without instances of CH. *Annals of Pure and Applied Logic* 26 (1): 75–87.

42. Shelah, Saharon. 1990. Universal graphs without instances of CH: revisited. *Israel Journal of Mathematics* 70 (1): 69–81.
43. Shelah, Saharon. 1994. *Cardinal Arithmetic*, volume 29 of *Oxford Logic Guides*. New York: The Clarendon Press Oxford University Press. Oxford Science Publications.
44. Silver, Jack. 1975. On the singular cardinals problem. In *Proceedings of the International Congress of Mathematicians (Vancouver, B. C., 1974)*, Vol. 1, pp. 265–268. Canad. Math. Congress, Montreal, Que.
45. Solovay, Robert M. 1965. $2^{\aleph_0}$ can be anything it ought to be. In *The theory of models. Proceedings of the 1963 International Symposium at Berkeley*, ed. by J. W. Addison, Leon Henkin, and Alfred Tarski, pp. 435, Amsterdam, North-Holland Publishing Co.
46. Solovay, Robert M. 1974. Strongly compact cardinals and the GCH. In *Proceedings of the Tarski Symposium, Proceedings of Symposia in Pure Mathematics*, ed. by Leon Henkin, John Addison, William Craig, Dana Scott, and Robert Vaught, Vol. XXV, pp. 365–372. Published for the Association for Symbolic Logic by the American Mathematical Society, Providence, R. I.
47. Walsh, S. 2012. Comparing Peano arithmetic, Basic Law V, and Hume's Principle. *Annals of Pure and Applied Logic* 163: 1679–1709.
48. Walsh, S., and S. Ebels-Duggan. 2015. Relative categoricity and abstraction principles. *The Review of Symbolic Logic* 8 (3): 572–606.
49. Wright, Crispin. 1983. *Frege's conception of numbers as objects*. Aberdeen: Aberdeen University Press.
50. Zermelo, Ernst. 1930. Über Grenzzahlen und Mengenbereiche. *Fundamenta Mathematicae* 16: 29–47.

Interpretation and Truth in Set Theory

Rodrigo A. Freire

Abstract The present paper is concerned with the presumed concrete or interpreted character of some axiom systems, notably axiom systems for usual set theory. A presentation of a concrete axiom system (set theory, for example) is accompanied with a conceptual component which, presumably, delimitates the subject matter of the system. In this paper, concrete axiom systems are understood in terms of a double-layer schema, containing the conceptual component as well as the deductive component, corresponding to the first layer and to the second layer, respectively. The conceptual component is identified with a criterion given by directive principles. Two lists of directive principles for set theory are given, and the two double-layer pictures of set theory that emerged from these lists are analyzed. Particular attention is paid to set-theoretic truth and the fixation of truth-values in each double-layer picture. The semantic commitments of both proposals are also compared, and distinguished from the usual notion of ontological commitment, which does not apply. The approach presented here to the problem of concrete axiom systems can be applied to other mathematical theories with interesting results. The case of elementary arithmetic is mentioned in passing.

1 Introduction

A foundational analysis of meaning and truth in axiomatic theories usually begins with a division of axiom systems in two groups: An axiom system is said to be *concrete* if its language is supposed to be interpreted in a specific way. In opposition, an axiom system is said to be *abstract* if it is not supposed to be interpreted in a specific way. For example, when explaining the construction of axiom systems in mathematics, Shoenfield writes:

> We have so far supposed that we have definite concepts in mind. Even so, it may be possible to discover other concepts which make the axioms true. In this case, all the theorems proved will also be true for these new concepts. This has led mathematicians to frame axiom systems

R. A. Freire (✉)
Philosophy Department, University of Brasília, Brasília, Brazil
e-mail: rodrigofreire@unb.br

W. Carnielli and J. Malinowski (eds.), *Contradictions, from Consistency to Inconsistency*, Trends in Logic 47, https://doi.org/10.1007/978-3-319-98797-2_9

> in which the axioms are true for a large number of concepts. A typical example is the set of axioms for a group. We call such axiom systems *modern* axiom systems, as opposed to the *classical* axiom systems discussed above. Of course, the difference is not really in the axiom system, but in the intentions of the framer of the systems. ([12], p. 2)[1]

I agree with Shoenfield that it is unproblematic that the difference is not really in the systems. However, how could one fix the system's intended interpretation? There are obvious problems with the claim that the intentions of the framer of the system already fix the interpretation. For it is not enough to intend. It is not clear how intentions could fix an interpretation, for the intention to present an axiom system about some specific subject matter may not suffice to interpret the axiom system.

When we are introduced to axioms for (first-order) arithmetic we are (usually) supposed to learn that they are about definite concepts of sum and product of natural numbers. But, using Shoenfield's formulation, there are other concepts which makes the usual axioms of arithmetic true. An axiom system is constituted by sentences which are purely linguistic objects, and its interpretation, what it is intended to be about, is therefore a matter of language convention. The acquired meaning of linguistic symbols is given by some kind of convention. What is the language convention according to which it is possible to interpret axioms for arithmetic in the usually intended way? One such language convention must *fulfill* intentions, but it cannot be *made* of intentions. For it seems clear that we do not learn intentions, whatever this could mean, when we learn that an axiom system for arithmetic talks about the definite concepts of sum and product of natural numbers, because it is not clear how to constitute a language convention from intentions.

This paper is first of all concerned with the following

Problem 1: *How could we explicate the commitments latent in the interpretation of an axiom system for set theory we happen to have inherited?*

If an axiom system is supposed to talk about some definite concepts, then there must be something prior to the system in question that we have to adopt in order to interpret its language. Otherwise, how could the sentences of an axiom system acquire the meaning that fulfills the intention to talk about specific concepts? As I have argued above, although this prior standard that we have adopted when we learn to interpret an axiom system for set theory must fulfill the relevant intentions, it cannot be a plurality of intentions. Now, the relevant question is: What should we adopt in order to interpret an axiom system for set theory in some specific way?

Of course, one could say "why bother? If we do not know how to make sense of the notion of interpreted formal system, just forget about it. It does not make any difference when we are doing the technical work with formal systems." Unfortunately, the problem of making sense of the notion of interpreted formal system is directly related to the problem of fixing truth-values. Based on Gödel's incompleteness theorems, arithmetical truth is generally held to go beyond provability, implying that the axiom systems for arithmetic must be regarded as concrete. Therefore, there must be an interpretation of the language of arithmetic fixing truth-values beyond provability.

---

[1] The terms *concrete* and *abstract* are used, for example, by Tait in [13], p. 90.

Making sense of the notion of interpreted formal system is a required component in any explanation of this transcendent nature of arithmetical truth.

Notice that problem 1 is only asking how can we understand the notion of concrete axiom system for set theory, that is, how can we explicate interpreted character of the language of set theory that excludes from consideration at least some concept of set membership which nevertheless makes the axioms true. The thesis that there is one correct concrete axiom system for set theory is a substantive thesis that I will discuss later.

In general, we say that a *sentence* is true with respect to a *prior standard* if an *agreement* between sentence and standard obtains. Otherwise, we say that the sentence is false with respect to the instituted standard. The commitments latent in an interpretation of an axiom system for set theory amounts to such a prior standard.

There is a traditional position on this problem according to which an interpreted axiom system for set theory is an axiom system endowed with a model, or a class of models which constitute a prior standard for the correctness of the axiom system, existing independently of our mathematical practice. These *independent models* cannot be the usual mathematical models. In fact, usual mathematical models are mathematical objects living inside set theory, and cannot be used to define the presumed concrete character of set theory. In this picture, the interpretation of an axiom system for set theory is committed to the existence of independent models.

Furthermore, in this framework, we are assumed to have some kind of "nonpropositional grasp", or "mathematical intuition" of independent models, from which we interpret the axiom systems and ground the axioms. This view will be discussed later. For now, it is enough to say that I think this position is inadequate. Models for set theory are very complicated, there is no simple representation of such a model, and it is not reasonable to say that models for set theory are, historically or conceptually, prior to the axioms. It is not clear how we could have some kind of "nonpropositional grasp" of those models, and how we could extract a language convention connecting the axiom system and its models from it.

I will work out a completely different solution to problem 1 based on the following thesis: If an axiom system is supposed to be interpreted in a specific way, then there must be a public *criterion*, given as principles which are considered unambiguous, objective and based on the standard practice of the relevant mathematical theory[2] and prescribe the interpretation of the system, that is, the intended class of models. What do we learn when we learn to interpret an axiom system for set theory is a list of principles that prescribe that interpretation. It is not necessary to assume the existence of standard models independent of our mathematical practice in order to *prescribe*, through a criterion, what would count as one.

In Sect. 2, two lists of principles corresponding to two criteria for interpreting the axiom systems for set theory will be explicitly stated, and these criteria constitute two

[2]The standard practice of set theory, that is the mathematical activity initiated by Cantor's seminal works, is *historically given*. The historical existence of this practice, considered as a merely historical phenomenon, is, of course, independent of the set-theoretic principles extracted from it. Therefore, there is no circularity in the relation between the set-theoretic principles that will be given here and the practice of set theory.

solutions for problem 1. The principles in those lists will be called *directive principles*. I assume, therefore, the priority of a list of directive principles over the corresponding class of standard models, and the axiomatization of a mathematical theory must be guided by such a list. The primary standard for the correctness of a mathematical theory is the corresponding list of directive principles, the corresponding class of standard models must be understood as a secondary standard for the correctness of the theory, and only insofar as it represents the principles. Each list of principles constitutes a direction to be pursued by the theory.

After stating explicitly the directive principles for the two lists, I will be concerned with the following:

Problem 2: *How and to what extent does the interpretation of the axiom systems for set theory committed to the criterion corresponding to given directive principles fix truth-values?*

In Sect. 3, an analysis of problem 2 will be given.

The analysis of problems 1 and 2 is independent of the substantive thesis according to which there is one correct interpretation of the axiom systems for set theory. However, it will be argued that the interpretation prescribed by each list of directive principles stated below is a plausible understanding of the axiom systems for set theory. Its plausibility ultimately comes from the fact that it is an interpretation prescribed by principles extracted from the *standard* practice of set theory. Therefore, the solutions to problem 1 given in this paper are at least historically correct in the sense that it is based on the thesis that the presumed concreteness (or classicality) of axiom systems for set theory must rest on instructions by which we understand sets according to Cantor, Dedekind, Fraenkel, Hilbert, Zermelo, etc. I defend the thesis that the organization of set theory is not ultimately based on, or committed to, a nonpropositional grasp of independent objects, nor on intentions, but on a criterion given as a list of set-theoretic principles extracted from the standard practice of this mathematical theory.

Vann McGee has written a paper on a related subject [10]. In his paper's introduction McGee says:

> The internal problem is this: the realist conception supposes that the meaning of mathematical terms is fixed with sufficient precision to ensure that each sentence has a determinate truth value. Now whatever meaning a linguistic expression has it possesses in virtue of the thoughts and practices of human beings. Not all meaning is thus dependent on human thought and action – the fact that a red sky in the morning means stormy weather isn't a matter of convention – but the fact that the numeral '7' refers to the fourth prime is a matter of how we have chosen to use a symbol. So there must be something we think, do, or say that fixes the intended meaning of mathematical terms. How are we able to do this? Mathematical objects aren't like Fido, whom you can hold by the collar while you give him a name. Nonetheless, there is something we think, do, or say that connects concrete speech acts with their abstract referents. Until we can give at least a rudimentary account of how this is done, the realist doctrine that mathematical sentences have determinate truth values will remain deeply mysterious. ([10], pp, 35–36)

Although closely related, McGee's internal problem of realism and problem 1 on axiom systems for set theory are, strictly speaking, different problems. They are conceptually different, and there are at least two important practical differences between McGee's problem and problem 1. First, McGee's problem presupposes that each sentence has a rigid truth-value. This is not required for an axiom system to be concrete. In fact, an axiom system is not to be considered concrete if it is understood as talking about whatever structure satisfying the axioms; otherwise it is interpreted according to a criterion. Thus, to say that an axiom system is interpreted in a specific way is not to say that its subject matter is a unique structure up to isomorphism, or a class of structures which are all elementary equivalent, but that it is not the class of *all* structures which satisfy the axioms. Secondly, for both McGee's problem and problem 1, the concrete character of an axiom system must rest on mathematical practice, but for a criterion to be considered a solution to problem 1 it must prescribe a class of structures and be extracted from the *standard* practice of set theory. The latter condition is not required for a solution of McGee's problem. McGee's proposal is stated in general outline in the following paragraph:

> Knowing how to use mathematical terms in practical problem solving is an important component of our understanding of mathematical vocabulary, but it doesn't take us far enough. Something more is needed. The format of our proposed answer is this: What we learn when we learn mathematical vocabulary, apart from learning how to count and measure, is a body of mathematical theory. What else could the answer be? The meaning given to a mathematical term is wholly dependent upon our use of the term (unlike a term like "Fido", whose meaning depends partly on our usage and partly on causal connections beyond our control), and our practical uses of the term aren't enough to determine the truth values; so what else is left but our use of the term in theorizing? ([10], p. 40)

There are important similarities and dissimilarities between McGee's solution to his problem and the solutions to problem 1 that I will present in this paper, and the comparison of these solutions is illuminating. I am arguing that the solution to problem 1 is that a concrete axiom system is a formal system accompanied by a criterion given as a list of directive principles which gives the commitments latent in the interpretation of set theory and, although these principles must be given as set-theoretic principles extracted from standard practice of set theory, they are *not* a separate mathematical theory as in McGee's proposal. Another point is that McGee's solution is not faithful with respect to the standard practice of set theory. Indeed, McGee appeals to an *Urelement Set Axiom* ([10], p. 52), and this axiom is not to be found in the standard practice of set theory. On the other hand, the solution to problem 1 that will be presented shortly is not a solution to McGee's problem, because it does not fix all truth values. There would be more interesting things to say in this regard, but I shall not be occupied with an exegesis of McGee's paper.

2 Two Double-Layer Pictures for Set Theory

As I have already said in Sect. 1, I propose that a criterion, given by set-theoretic principles, must be the basis for a solution of problem 1, and these principles will be called directive principles. From now on, for the sake of definiteness, I will identify axiom systems with formal first-order systems, in the usual sense. The solution that I propose is roughly as follows: Formal systems for set theory can be considered as concrete if and only if they are, tacitly or not, accompanied by a criterion, given as a list of directive principles, which prescribe its interpretation, that is, the appropriate class of structures. There is a list of minimal *desiderata* that a proposed solution to problem 1 has to meet in order to be treated as a plausible understanding of the axiom systems for set theory: (i) A solution to problem 1 is based on set-theoretic principles, the directive principles, that must be explicitly stated. (ii) All these principles must be extracted from the standard mathematical practice of set theory, according to the thesis that, whatever meaning a linguistic expression in this theory has, it has in virtue of the standard practice of set theory. (iii) It is desirable that the instructions contained in the directive principles solving problem 1 relate to the formal systems for set theory in an unproblematic way, without appeal to any intuition of independent objects. (iv) One such solution must fix truth-values in a satisfactory way, that is, at least all arithmetical statements must have rigid truth-values in the interpretation given by the corresponding criterion.

Of course, since the directive principles are stated explicitly below, it follows that this solution meets *desideratum* (i). The following directive principles highlight the *production* of sets, which is undoubtedly a central aspect of set theory.

2.1 The First List of Directive Principles

1. A set is determined by its elements, which are sets themselves, and there is no infinite regress in this transitive determination.
2. An arbitrary choice of elements of a set determines a set, which is a subset of the original set.
3. An arbitrary replacement of each element of a set by a set determines a set.
4. All the elements of the elements of a set determine a set.
5. All the subsets of a set determine a set.
6. There is an infinite set.

The first thing to say about the criterion given by the directive principles above is that they are *not* an independent formal system for set theory, nor a natural language formulation of the axioms of a formal system. They are not even a separate theory but just give a criterion for interpreting the formal systems for set theory. This criterion *is not defining sets*, in any significant sense. Recall that, for the language of set theory to be understood as being about something more specific than whatever concept of set membership which makes the axioms true, there must be some prior standard, a criterion excluding some of these concepts, that we learn when we learn to interpret it. The directive principles can play the role of this prior standard: They are the basic directives for an axiomatic theory of sets that can be found in the main line of thought developed throughout set theory, and this is the only a priori justification needed for them, as they are not, in any philosophically significant sense, "absolute set-theoretic principles".[3]

Now, notice that the formal system ZFC can be obtained from the directive principles by formalization. For example, each instance of the replacement axiom is directly obtained from principle (3): Given a set A and a functional formula F, principle (3) gives the unique set B obtained from A by replacing each element by its F-image, which is the set required to exist by the replacement axiom. The only nontrivial case is the axiom of choice, but this axiom can also be obtained by a formalization of the principles: A choice set for a given set is a subset of the union of the set, and hence can be obtained by the use of directive principles (2) and (4). It can also be easily obtained from directive principle (3) alone, as the production of a choice set for a given set is the particular case of a replacement of each element of the given set by an element of itself.[4] The axiom of choice can be seen as an attempt at formalizing part of the *arbitrariness* present in directive principles (2) and (3). In Sect. 3 I will show that this arbitrariness cannot be fully formalized.

In order to make explicit the point that the criterion given by the directive principles above meets *desideratum* (ii), that is, that they are historically coherent with the formal systems for set theory, I will say a few words on the origins of the principles: Harward and Cantor, although *not working from an axiomatic perspective*, discussed what they considered to be very basic facts about sets. The former carried forward his discussion in an article at the Philosophical Magazine in 1905, and the latter in correspondence with Dedekind and Hilbert. The facts that they noted are very similar to the directive principles stated above: Both suggested something like principles (2) to (6). For example, Cantor in a letter to Dedekind ([2], p. 114) stated the following:

> Two equivalent multiplicities either are both "sets" or are both inconsistent.
> Every submultiplicity of a set is a set.
> Whenever we have a set of sets, the elements of these sets again form a set.

Directive principles (2), (3) and (4) are *collectively equivalent* to these three statements in Cantor's letter. Of course, principles (5) and (6) are also a basic part of Cantor's seminal works.

Principle (1) is the outcome of the guiding thought according to which the right way to think about a set is that it is, whatever its nature may be, a well-founded, extensional entity determined directly by its members. This thought can be reformulated as [A set is an object whose immediate constituents are its elements, and it is determined by its elements. Furthermore, since an object cannot be a constituent, immediate or

---

[3]This is an important point: The directive principles above give instructions for understanding formal sentences in usual set theory – these principles are *not* directly related to other set theories such as those theories of non-well-founded sets. Therefore, their application outside the framework of usual set theory is not justified.

[4]This is not accidental. In fact, with the exception of the empty set, which can be produced by directive principle (2) from any given set, all subsets produced by directive principle (2) from a given set can also be produced by directive principle (3). Indeed, it is easy to see that for each arbitrary choice of elements of a given set producing a nonempty set there is a replacement of the elements of the given set producing the same nonempty set.

not, of itself, it follows that sets should be extensional and well-founded objects.] The following famous passage of [3] is relevant for this thought and, consequently, for directive principle (1):

> By an "aggregate" (*Menge*) we are to understand any collection into a whole (*Zusammenfassung zu einen Ganzen*) M of definite and separate objects m of our intuition or our thought. These objects are called the "elements" of M. ([3], p. 85)

It seems clear from this passage that, according to Cantor, a set is determined by its elements.[5] The extensionality of sets is surely a component of the set theory developed by Cantor, but it seems that he nowhere explicitly talks about it. Dedekind states it as a fact about sets in his famous 'The Nature and Meaning of Numbers'. An explicit thought on the well-foundedness of sets is to be found on early drafts of Zermelo's axioms (preceding the publication of [14]) where, according to Moore ([11], p. 165), he had initially assumed, as a postulate, that a set could not be a member of itself.

I will not make any further comments on the origin of the directive principles in the given list. The above paragraphs are meant as a brief illustration of how these principles can be seen as the outcome of guiding thoughts on sets by the founding fathers of set theory, sometimes working in different perspectives, with different intentions and independently of each other. The directive principles are a primary expression of directions that led to the formal systems for set theory, and, as I understand it, one such expression must be part and parcel of this theory. This account for the possibility of interpretation of formal systems for set theory is *not* a theory about the psychological origin of the axioms of those formal systems. Accordingly, directive principles are not intentions, intuitions or dialectical considerations about sets; they constitute a *public* criterion for interpreting formal sentences in set theory and this historical note is intended to show their correctness with respect to the standard practice of this mathematical theory.

The codification of the directive principles in a formal (first-order) system give rise to nonlogical axioms in a formal language, but the principles *precede* their formal counterparts in the schema of set theory. The formal systems for set theory can be obtained from the directive principles by an operation of formalization. Since formalization does not appeal to any intuition of mathematical objects, the proposed solution meets *desideratum* (iii). Also, it must be obvious that we should not throw the directive principles away after formalizing set theory, unless we want to give up the possibility of interpreting the resulting formal systems according to those principles. Thus, I am proposing that set theory consists of two layers: The directive principles are the first layer, and the formal counterparts of the directive principles are the second layer. As it is already known, the formal system ZFC is one such formal

---

[5]It is also important to notice that in this passage Cantor seems to express the view according to which all sets are psychological objects. This is *irrelevant* for understanding a theory of sets: There may be concrete and abstract, psychological and physical sets, and a theory of sets must only be concerned with aspects common to all these possibilities. Therefore, the only important point extracted from this passage is that a set is, *whatever its specific nature may be*, determined by its elements without any further addition.

counterpart of the directive principles, which means that ZFC is in the second layer in this double-layer schema. Also, the arbitrariness present in directive principles (2) and (3) cannot be fully formalized, which means that the second layer in the double-layer schema is open-ended.

Furthermore, reinforcing what was said above, although the directive principles above can be equivalently written in the second-order language of set theory, they are *not* a kind of "second-order formal system". The directive principles do not constitute any kind of autonomous formal system in the sense that they do not play a deductive role in the double layer schema. Therefore, there is emphatically no attempt to replace the usual (first-order) formal systems for set theory by directive principles – this would be a very bad move because these principles cannot play the deductive role of a formal system. The double-layer schema of set theory stated above is based on the complementarity of directive principles and formal systems: The principles demarcate the subject matter of the formal systems for set theory by giving a criterion for excluding at least some concepts of set membership which nevertheless makes the axioms true, because they constitute stronger commitments, but they cannot replace their formal counterparts in proofs.

2.2 The Second List of Directive Principles

1. A set is determined by its elements, which are sets themselves, and there is no infinite regress in this transitive determination.
2. The elements of a set satisfying a property determines a set, which is a subset of the original set.
3. The replacement of each element of a set by a set given by a functional relation determines a set.
4. All the elements of the elements of a set determine a set.
5. All the subsets of a set determine a set.
6. Given a countable list of sets, there is a set determined by exactly those sets in the list.
7. Given a set there is a choice function on this set.

The general remarks about directive principles given above apply equally to this second list. Now, in contrast with the first list, sets given by replacement and separation are supposed to be *defined*. There is a stronger principle of infinity, principle (6), which can also be understood as an infinitary replacement axiom for countable sets, for it can be reformulated as follows: An arbitrary replacement of each element of a countable set by a set determines a set. One traditional argument for the replacement axiom states that without this axiom it is not possible to prove the existence of the set $\{\omega, \wp(\omega), \wp(\wp(\omega)), ...\}$. Of course, we can argue that there is nothing special with this countable set, and this is an evidence that the stronger principle of infinity is instituted in the practice of set theory, used to justify the replacement axiom. In fact, we can find the stronger principle of infinity expressed in [6], p. 46, footnote 5 (and also in [1]):

> A stronger axiom schema of infinity than **VI** is introduced in Fraenkel 27 (p. 114, Axiom **VII**c. Fraenkel's axiom is equivalent, on the basis of axioms **I**-**V**, to the schema which asserts, roughly, that every "denumerable collection of elements" is a set (Bernays 37–54 **III**).

Furthermore, the notion of countable list of sets is present in set theory from its beginnings. For example, Cohen says ([4], p. 64):

> It is perhaps not generally known, but Cantor's stimulus to study set theory arose from countable ordinals.

A choice principle concludes our list, which, of course, is extracted from the standard practice of the discipline.

A first-order (partial) formalization of the above principles results in ZFC, naturally. These principles can be equivalently written in the infinitary language $L_{\omega_1\omega_1}$. The resulting list of sentences is equivalent to ZFC plus

$$\forall x_0 \forall x_1 ... \exists y \forall z (z \in y \leftrightarrow z = x_1 \vee z = x_2 \vee ...).$$

This second list of principles has the virtue that it can be equivalently written in the language $L_{\omega_1\omega_1}$, while the first list requires the much more complex second-order language. In other words, the meaningfulness of the criterion given by the second list of principles is a cheap assumption when compared to the meaningfulness of the criterion given by the first list. Unfortunately, based on a mathematical analysis of the proposed lists of directive principles, it is arguable that only the first list captures the conception of *cumulative hierarchy of sets*, which seems to be the basic concept in terms of which set-theoretic truth is usually understood. The analysis of both lists that will be given in Sect. 3 clarifies the differences in a precise way.

The first layer in each of these double-layer pictures gives a criterion for the interpretation of formal systems for set theory; the second layer is always the domain of proof and formalization within the finitary range. The double-layer pictures of set theory that I have proposed have nothing to do with the two layers of mathematics in the picture of Platonism as diagnosed by Tait in the criticisms of Benacerraf and Dummett:

> Benacerraf and Dummett seem to me to be typical of those who adopt a particular picture of Platonism. The picture seems to be that mathematical practice takes place in an object language. But this practice needs to be explained. In other words, the object language has to be interpreted. The Platonist's way to interpret it is by Tarski's truth definition, which interprets it as being about a model – a Model-in-the-Sky – which somehow exists independently of our mathematical practice and serves to adjudicate its correctness. So there are two layers of mathematics: the layer of ordinary mathematical practice in which we prove propositions such as [There is a prime number greater than 10] and the layer of the Model at which [There is a prime number greater than 10] asserts the 'real existence' of a number. ([13], p. 67)

Later on, Tait concludes

> The myth of the Model tends to get attached to Platonism (or at least to 'epistemological' Platonism in the sense of Steiner (1975)) because the view that mathematics is about things

> like the system of numbers is compared with the view that propositions about sensible things are about the physical world; and here there is a tendency to believe that there *is* such a nonpropositional grasp, namely, sense perception, which does endow meaning on what we say and to which we appeal to determine truth. But I hope that, if not what I have said, then Wittgenstein's critique of this view of discourse about sensibles will convince the reader that it is inadequate. ([13], p. 74)

There is no such thing as the layer of the "Model(s)-in-the-Sky" in the double-layer accounts of set theory given here, and there is no appeal to a nonpropositional grasp of independent objects. First of all, the double-layer accounts given here are not committed to independent objects, but are only committed to the objectivity of the criteria given by directive principles. Furthermore, there is no need of any kind of nonpropositional grasp, which is usually needed in a schema in which the possibility of interpretation is alienated from formal systems. If the second layer of a double-layer schema consists of first-order counterparts of the first layer, as is the case here, then the formal system's corresponding interpretation is not alienated from them and there is no nonpropositional grasp involved.

With respect to the particular picture of Platonism which seems to be adopted by Benacerraf and Dummett in their criticisms, Tait writes

> ... Needless to say, it is not this version of Platonism that I am defending or that I even understand. Thus, I should not be understood to be taking part in any realism/antirealism dispute, since I do not understand the ground on which such disputes take place. As a mathematical statement, the assertion that numbers exist is a triviality. What does it mean to regard it as a statement outside of mathematics? ([13], p. 68)

The directive principles are not subject to Tait's question, because they are neither statements outside set theory, nor ways of stepping outside this mathematical framework. These principles just give us a criterion for the interpretation of formal systems for set theory. When mathematically analyzed, each list of directive principles gives a criterion separating a class of *standard* structures and giving the truth conditions for each sentence, as it will be shown in Sect. 3.

2.3 Other Views on Set Theory and Mathematical Truth

A possible alternative to a double-layer schema of set theory is, of course, a single-layer account of it, according to which set theory consists only of a layer of first-order formal systems, which can be either static or evolving in time, such that their formal languages are not supposed to be about something more specific than all possible models of their axioms. It seems to me that this single-layer picture is inadequate in several ways. I will mention three: First, it is historically problematic, because the original framers of set theory did not think about set theory that way, as an abstract axiom system, and, for sure, the original stimulus to study set theory was not to encompass a plethora of concepts of set membership. Secondly, it can be seen as an artificial attempt at reducing truth to provability by decree, while at the same time saying nothing about the justification of the axioms. Thirdly, in this single-layer view of set theory, we have no way to understand $\in$ as capturing a concept of membership: It is just an arbitrary relation which satisfies a given list of axioms of a formal system for set theory. In this case, not even the interpretation of finitary sentences is satisfactory, for different models, possibly non-wellfounded, may disagree with respect to these sentences. However, it is very reasonable to assume that we can interpret finitary sentences in a satisfactory way, for, otherwise, it would be very hard to see how we could understand the formal systems themselves. More dramatically, from Gödel's second incompleteness theorem it follows that the consistency of set theory itself is one of those finitary sentences that witnesses this shortcoming of interpretation in the single-layer account of set theory in question. Gödel himself remarked that:

> ... It is *this* theorem [the second incompleteness theorem] which makes the incompletability of mathematics particularly evident. For, *it makes it impossible that someone should set up a certain well-defined system of axioms and rules and consistently make the following assertion about it: All of these axioms and rules I perceive (with mathematical certitude) to be correct, and moreover I believe that they contain all of mathematics.* If somebody makes such a statement he contradicts himself. For if he perceives the axioms under consideration to be correct, he also perceives (with the same certainty) that they are consistent. Hence he has a mathematical insight not derivable from his axioms. ([9], p. 309)

Although I do not believe that mathematical knowledge can be extracted from perception, I agree with Gödel that the consistency of a mathematical theory is a presumed component of the thought that led to its formal systems. Therefore, I argue that a satisfactory account of the truth of the axioms of a formal system for a mathematical theory must also account for the truth of their consistency, if they are consistent, and, at the same time, leave some room for the possibility of inconsistency. In fact, our grasp of formal systems *presupposes* that we understand mathematical sentences of the form "S is provable from the axioms" either affirmed or denied. This is accomplished in the double-layer pictures of set theory that I am proposing. For, on the one hand, if the formal systems for set theory are consistent then the truth value of their consistency is fixed by both the first and the second lists of directive principles, as it was seen in Sect. 3. On the other hand, if the formal systems for set theory are inconsistent, then both lists of directive principles are *eo ipso* incoherent and fail to play any role as a criterion for the interpretation of the formal systems.

This possibility of failure is an important aspect of this account of set theory because, indeed, we may fail. How could we fail to formulate a consistent system of axioms and rules if we do have a nonpropositional grasp of a "Model-in-the-Sky", which is declared as the standard for mathematical correctness? How could our formal systems for set theory turn out to be inconsistent if we *perceive with mathematical certitude* the truth of their axioms? The directive principles do not give us certainty regarding consistency: They can only give us a criterion for interpreting the formal language such that if the consistency sentences are *assumed to be true*, then they are true according to the given interpretation, which is exactly what provability cannot accomplish according to Gödel's second incompleteness theorem. The delimitative role played by directive principles when demarcating the class of standard models of the theory and fixing truth-values is very different from the role of "mathematical intuition" as a prior standard that the axioms have to meet in order to be considered correct. In [7], Gödel argued for "mathematical intuition" as a criterion of truth in set theory different from provability in a formal setting:

> ... What, however, perhaps more than anything else, justifies the acceptance of this criterion of truth in set theory is the fact that continued appeals to mathematical intuition are necessary not only for obtaining unambiguous answers to the questions of transfinite set theory, but also for the solution of the problems of finitary number theory (of the type of Goldbach's conjecture), where the meaningfulness and unambiguity of the concepts entering into them can hardly be doubted. This follows from the fact that for every axiomatic system there are infinitely many undecidable propositions of this type. ([7], p. 485)

I do not think that an appeal to some obscure mathematical intuition clarifies anything. I think that a priori mathematical intuition is something we know nothing about. We know much more about mathematical truth than about a priori intuition: explaining the former in terms of the latter sounds like explaining the partially understood in terms of the completely not understood. If mathematical intuition is understood as a posteriori mathematical feeling, then it is something we *gain* through continued mathematical training. We could be trained in a mathematical subject until the basic articulation of its primitive notions become fully grasped and we forget how it was when we were not yet thinking that way. At this point, we could say that the fundamentals of the subject became intuitive to us, but to say that this a posteriori intuition is the foundation of the subject is to reverse the order of explanation. I am defending a replacement of 'mathematical intuition' by 'directive principles', which, of course, is not a change of names – their roles and nature are very different. With this proviso, I would agree with a thesis that is related to Gödel's extract above: As soon as we want to determine in what relation set theory stands to and to what extent it is captured by formal systems, then analysis of directive principles is essential. The reason behind this is that formal systems cannot provide the criteria for structures that directive principles can.

Trying to clarify what is meant by mathematical intuition, Gödel remarked earlier that:

> It should be noted that mathematical intuition need not be conceived of as a faculty giving an *immediate* knowledge of the objects concerned. Rather it seems that, as in the case of physical experience, we *form* our ideas also of those objects on the basis of something else which *is* immediately given. Only this something else here is *not*, or not primarily, the sensations. ([7], p. 484)

Gödel does not tell us what this "something else which is immediately given" is supposed to be, but there are some important remarks on the directive principles that are related to Gödel's attempt to clarify the issue of mathematical intuition: The directive principles are not a faculty giving an *immediate* knowledge of sets. Rather it seems that we form our ideas of the mathematical objects concerned on the basis of the directive principles in the sense that the subject matter of set theory is prescribed by these principles. They are an outcome of the thoughts of the founding fathers of this mathematical theory, but their origin is not a primary concern of the investigations on the possibility of interpreting formal systems for set theory. In any case, it is clear that they are not sensations.

Ferreirós ([5], pp. 380–384) is another logician-philosopher that has recently expressed a related view on the insufficiencies and implausibility of an account of mathematical theories according to which they consist of formal systems standing alone. When talking about two contrary tendencies in the development of mathematics, one which aims at a reduction of mathematics to a purely symbolic system and another one which aims at a reduction of mathematics to a purely conceptual system, Ferreirós writes

> In my view, the failure of both radical tendencies is of the essence. The standpoint I adopt emphasizes the need to consider the meaning or thought that accompanies formulae and calculations. (This is no doubt shared by many other philosophers, but the question is how to proceed.) Mathematical symbolism cannot be mastered without immersion in a practice, and by learning the practice we learn to associate representations and meaning to the formulae. Normal (so-called informal) symbolic systems and theories cannot be made to stand alone outside of practice; and when systems and theories are formalized and made to stand alone, the phenomenon of non-standard interpretations arises in a natural way.
>
> Indeed, I defend the *complementarity of symbolic means and thought* in mathematics–each one joined by the other, none of them reducible to the other. For obvious reasons, it is more difficult to deny the role and importance of the symbolic component in mathematics, but substantial arguments can be given for a similar conclusion concerning the conceptual component. For my purposes here, I shall be content with the modest claim that, in light of developments in mathematics and its foundations during the 20th century, such a standpoint deserves to be seriously considered as an option. ([5], pp. 381 and 382)

Of course, I agree with Ferreirós on the insufficiency of symbolic means to account for (all) mathematical theories. However, it is not enough to say that mathematics and set theory are a combination of a symbolic component with a conceptual component – if one is trying to formulate in these terms a standpoint on the foundations of mathematical truth then one must say what the symbolic and conceptual components are supposed to be and how they relate to each other. I am defending that, in the case of set theory, *directive principles*, which give unambiguous and objective criteria to demarcate the interpretation and understanding of formal systems, constitute the conceptual component[6] complementing the layer of formal systems, and not *thought*, *meaning* and *representations*, which are the categories used by Ferreirós in the above extract. Ferreirós does not tell us what exactly this conceptual component he is talking about is: It is not even clear whether this conceptual component is public or private. I do not think that the intentions, representations and thoughts that we associate to formulae when we learn a practice are univocal. The problem of the possibility of interpreting formal systems is in need for a more precise account of the conceptual component of the corresponding concrete axiom systems.

I am proposing the complementarity of formal systems and the directive principles – formal systems cannot play the delimitative role that the directive principles can, and these principles cannot play the deductive role of formal systems.

[6] The conception of the conceptual as a criterion is unproblematic: A concept naturally gives rise to a criterion separating those things falling under it from the rest.

3 Mathematical Analysis of the Role of Directive Principles

Although lists of directive principles do not constitute a formal system, their role as criteria for the interpretation of formal systems can be *analyzed* within mathematics. Before someone says this is circular, I reinforce that I am not talking about any *reduction* of directive principles to formal systems. Circularity comes in only if a reduction is proposed: It is circular to reduce directive principles to formal systems, since those principles are a basis for interpreting formal systems. However, it is *not* circular to use a formal system for set theory to *analyze* directive principles. There is no reduction taking place here, the layer of the formal systems does not subsume the layer of the directive principles and, consequently, there is no circularity.

From now on, I will assume that the formal system ZFC for set theory is consistent. The aim of a mathematical analysis of the role of directive principles is to clarify what is, according to each of the two lists of principles, the subject matter of set theory.

3.1 The Standard Models of Set Theory According to the First List of Directive Principles

I take it for granted that a mathematical analysis of the subject matter of an axiom system for a mathematical theory must result in a distinguished class of structures, the standard models. Therefore, the following definition is required:

Definition 1 A structure (M, E), in which E is a binary relation on M, is said to *conform* to the first list of directive principles iff

1. (M, E) satisfies the axioms of extensionality and regularity.
2. If $x \in M$ and if c is a subset of M such that for each $y \in c$ it holds that yEx, then there is an element in M whose E-members are the elements of c.
3. If $x \in M$ and if $r : M \to M$ is a function on M, then there is an element in M whose E-members are all those elements w such that $r(y) = w$ for some y that is E-member of x.
4. If $x \in M$ then there is an element in M whose E-members are all E-members of E-members of x.
5. If $x \in M$ then there is an element in M whose E-members are all E-subsets of x.
6. (M, E) satisfies the axiom of infinity.

Theorem 1 characterizes those structures conforming to the first list of directive principles, clarifying what is, according to those principles, the subject matter of set theory:

Theorem 1 (Zermelo [15]) *A structure (M, E) conforms to the first list of directive principles iff for some strongly inaccessible cardinal κ, $(M, E) \cong (V_\kappa, \in)$.*

Proof First, notice that E is a well-founded relation on M. In fact, suppose that there is an infinite sequence $(x_i)_{i\in\omega}$ of elements of M such that $x_{i+1}Ex_i$, for each $i \in \omega$. Since (M, E) satisfies all the axioms of ZFC, it follows that every element in M has a transitive closure in the sense of (M, E). Let x be the transitive closure of x_0 in (M, E), and let c be the set $\{x_i : i \in \omega\}$. From the second directive principle in the first list, it follows that there is a set $y \in M$ such that zEy iff there is an $i \in \omega$ such that $z = x_i$. Thus,

$$(M, E) \models \forall z(zEy \rightarrow (\exists w(wEy \wedge wEz))),$$

and (M, E) cannot satisfy the axiom of regularity. Therefore, it is not the case that there is an infinite sequence $(x_i)_{i\in\omega}$ of elements of M such that $x_{i+1}Ex_i$, for each $i \in \omega$.

Now, let $(N, \in)$ be the transitive collapse of (M, E). If α is the first ordinal which is not in N, then $N \subseteq V_\alpha$. Suppose $N \neq V_\alpha$. The ordinal α is a limit ordinal. For if α is $\beta + 1$, then $\beta \in N$ and the set $\beta \cup \{\beta\} \in V_{\alpha+1} \setminus V_\alpha$ cannot be in N. From this it follows that $(N, \in)$ cannot be a transitive model of ZFC and we conclude that α is limit.

Let β be the first ordinal such that there is an element x in $V_\beta \setminus N$. Since α is limit, $V_\alpha = \bigcup_{\gamma<\alpha} V_\gamma$, and if for all $\gamma < \alpha$ it holds that $V_\gamma \subseteq N$, then $V_\alpha \subseteq N$. Therefore $\beta < \alpha$ and $\beta \in N$. Since $x \subseteq N \cap V_\beta = (V_\beta)^N \in N$, it follows, from the second directive principle, that $x \in N$. Therefore, $N = V_\alpha$.

From the third and fifth directive principles, it can be proved that α is regular and strong limit, respectively. In fact, if β is an ordinal such that $\beta < \alpha$, then $\beta \in N$ and the direct image of a function $r : \beta \rightarrow N$ is an element of N. Since the union of an element in N is also in N, it follows that $r : \beta \rightarrow N$ cannot be cofinal in α. Also, if $\lambda < \alpha$ is a cardinal, then $\wp(\lambda) \in N$, the cardinal of $\wp(\lambda)$ in the sense of $(N, \in)$ is 2^λ, and $2^\lambda < \alpha$. This proves the result. □

From this mathematical result, it is plausible to say that, according to the first list of directive principles, set theory is about the concept of the cumulative hierarchy of sets, which is exemplified by the hierarchies of sets $(V_\kappa, \in)$, in which κ is a strongly inaccessible cardinal. Now, is this the correct interpretation of set theory? Are the hierarchies of sets the *true* subject matter of set theory? I think that the only sure answer to this question is that this interpretation is historically correct, in the sense that (*a*) the concept of the cumulative hierarchy of sets seems to be the basic concept in the discussions on set-theoretic truth, and (*b*) the directive principles in the first list are to be found as the basic facts about sets in the works of the founding fathers of this mathematical theory, and fixes truth-values in a plausible way, as it will be shown in the sequel. Therefore, although directive principles are not intentions, they seem to fulfill the intentions of the framers of the axiom systems for set theory, notably Cantor's and Zermelo's intentions.

Now it is possible to clarify how sentences in formal systems for set theory acquire a meaning under the background language convention given by the first list of directive principles. Adopting a model-theoretic perspective on this point, if φ is a sentence (in the first-order language with a symbol for equality and one binary predicate variable R for membership) then the *meaning* of φ can be defined, according to the background language convention given by the first list of directive principles, as the class function that assigns to each structure $(V_\kappa, \in)$, in which κ is a strongly inaccessible cardinal, the truth-value **T** if $(V_\kappa, \in) \models \varphi$, and the truth-value **F** otherwise. This extensional account of meaning is based on the thesis that a sentence expresses its truth-conditions with respect to a presumed background language convention. In this model-theoretic setting, the truth-conditions of a sentence φ in a formal system for set theory can be identified with structures $(V_\kappa, \in)$ in which φ is true. The truth-value of φ is said to be *fixed by the directive principles* if and only if φ has the same truth-value on every structure $(V_\kappa, \in)$, in which κ is a strongly inaccessible cardinal. It is already easy to see that the directive principles and the first-order axioms of the formal system ZFC have very different fixing powers.[7] If there is an explanation to this greater fixing powers of the directive principles, then it is that, contrary to the axioms of a formal system, the directive principles are *not* required to effectively generate the truths that are fixed by them. The role of the directive principles is not to computably generate truths, but just to give a criterion for the interpretation of formal systems.

Recall that, in a model-theoretic perspective, if φ is a sentence then the truth-value of φ is fixed by the background language convention given by the first list of directive principles iff φ has the same truth-value on every structure $(V_\kappa, \in)$, in which κ is a strongly inaccessible cardinal, that is, iff the class function which is the meaning of φ is constant. The structures (isomorphic to) $(V_\kappa, \in)$, in which κ is a strongly inaccessible cardinal will be called *Z-standard models* of ZFC. The continuum hypothesis, CH, for example, has the same truth-value on every Z-standard model[8] because all sets relevant to the truth or falsity of CH belong to a lower level V_α, in which α is countable. Therefore, the truth-value of CH is fixed by the directive principles. Also, the truth-value of arithmetical sentences – consistency statements, in particular – are fixed[9] by the first list of directive principles, and, of course, the truth of every theorem of ZFC is fixed by the first list of directive principles.

The mathematical analysis of the truth-value of some statements can be conditional. Consider the sentence ‘there is a strongly inaccessible cardinal’. If there is a strongly inaccessible cardinal, then this sentence does not have the same truth-value on every Z-standard model. If there are no strongly inaccessible cardinals then the falsity of the sentence ‘there is a strongly inaccessible cardinal’ is fixed by the first

[7] Naturally, the truth-value of φ is said to be fixed by the axioms of ZFC iff φ has the same truth-value on every model of ZFC.

[8] Notice that this does not mean that CH holds in every Z-standard model. It just means that it is true/false in one Z standard model iff it is true/false in all Z-standard models.

[9] From Gödel’s first incompleteness theorem it follows that the sentences that hold in all standard models cannot be effectively enumerated. This shows that the arbitrariness present in directive principles (2) and (3) of the first list cannot be fully formalized.

list of directive principles. Although this is not as informative as for CH, it gives some information: If 'there is a strongly inaccessible cardinal' is true then its truth is not fixed by the first list of directive principles, and if it is false then its falsity is fixed by the first list of directive principles. Similarly, the existence of a measurable cardinal can be analyzed conditionally: If 'there is a measurable cardinal' is true then its truth is not fixed by the first list of directive principles, and if it is false then its falsity is fixed by the first list of directive principles.

Another example of a conditional analysis is the following: If $V = L$ holds then it holds in every Z-standard model and its truth is fixed by the directive principles. On the other hand, if $V = L$ is false, then its falsity can be fixed or not by the directive principles depending on how it fails: If there is a non-constructible subset of ω then $V = L$ is false in every structure $(V_\kappa, \in)$ and its falsity is fixed by the first list of directive principles. If there is a strongly inaccessible cardinal and the rank of the least-ranked non-constructible sets is greater than the first strongly inaccessible cardinal, then the falsity of $V = L$ is not fixed by the first list of directive principles. Therefore, it seems that the first list of directive principles fix truth-values in a satisfactory way, meeting *desiderata* (iv). For example, the directive principles fix the truth-value of every arithmetic statement, and every statement whose truth or falsity depends only on an initial segment of the hierarchies of sets bounded by the first inaccessible. It is also plausible to say that the existence of an inaccessible cardinal, and related statements, does not have the truth-value fixed.

If the truth of a sentence assumed to be true is not fixed by the directive principles, then its adoption as a new axiom for set theory must be accompanied by an extension of directive principles, in case one wants to keep the possibility of interpreting the resulting formal systems. If the truth of a new axiom assumed to be true is not in the range of the directive principles stated above, then these principles are insufficient for interpreting the new axiom system. On the other hand, if the truth of a sentence assumed to be true is fixed by the directive principles then its adoption as a new axiom need not be accompanied by an extension of directive principles, but this does not mean that its adoption as a new axiom is *justified* in the picture of set theory that I am defending. Because, even in case the truth of a sentence assumed to be true is fixed by the directive principles, its adoption as a new axiom may cause a mismatch between the layer of the directive principles and the layer of formal systems. In fact, if $V = L$ is true then its truth is fixed by the directive principles, but $V = L$ cannot be obtained by a formalization of part of the directive principles, which implies that $ZF + V = L$ is not a formal system for set theory according to the double-layer picture of set theory presented here. Therefore, in this double-layer scheme, the mere conditional fact that if a sentence is true then its truth is fixed by the directive principles does *not* suffice to justify the adoption of the sentence as a new axiom. In this schema, it is also required that the resulting formal systems be obtained from formalizations of the directive principles.

3.2 The Standard Models of Set Theory According to the Second List of Directive Principles

The mathematical analysis of the criterion given by the second list of directive principles is based on the following:

Definition 2 A structure (M, E), in which E is a binary relation on M, is said to *conform* to the second list of directive principles iff

1. (M, E) satisfies the axioms of ZFC.
2. If c is a countable subset of M, then there is an element in M whose E-members are the elements of c.

It is easy to prove that a structure conforms to the second list of directive principles iff it is isomorphic to a transitive model of ZFC closed under countable subsets.

Theorem 2 *A structure* (M, E) *conforms to the second list of directive principles iff* (M, E) *is isomorphic to a transitive model of* ZFC *closed under countable subsets.*

Proof If (M, E) conforms to the second list of directive principles, then it is well-founded. In fact, suppose that there is an infinite sequence $(x_i)_{i \in \omega}$ of elements of M such that $x_{i+1} E x_i$, for each $i \in \omega$. From the second item in Definition 2, it follows that there is a set $y \in M$ such that zEy iff there is an $i \in \omega$ such that $z = x_i$. Therefore, (M, E) cannot satisfy the regularity axiom.

Now, let $(N, \in)$ be the transitive collapse of (M, E). The transitive structure $(N, \in)$ is isomorphic to (M, E), and it conforms to the second list of directive principles. It is a model of ZFC and it is closed under countable subsets.

The other direction is trivial. □

Since all transitive models of ZFC agree on arithmetic statements, the second list of directive principles fixes the truth-values for all arithmetical sentences, which is an important desideratum for a criterion for the interpretation of ZFC. The status of the continuum hypothesis is less determined. The power set of ω is absolute for transitive models which are closed under countable subsets. Also, $\aleph_1$ is absolute for those models. Therefore, if $\aleph_1$ is the cardinal of $\wp(\omega)$ in one transitive model closed under countable sets, then it is really the cardinal of $\wp(\omega)$. Equivalently, if the continuum hypothesis is false, then its falsehood is fixed by the second list of directive principles.

The notion of countable set is also absolute for transitive models of ZFC closed under countable subsets, and it seems sensible to say that the absoluteness of the notion of countability is part of the very conception of set theory.

In the proposed double-layer schema, the conceptual component of set theory is given in the first layer as a criterion separating some structures from the class of all models of the corresponding formal systems. Theorem 2 above shows that according to the second list of directive principles the subject matter of set theory can be understood as constituted by the transitive models closed under countable subsets, which can be called K-standard models. This is probably not the usual way the conceptual component of set theory is understood, but this account of set-theoretic truth has the merit of fixing the truth-values of arithmetic statements, and even those of second-order arithmetic, because of the absoluteness of $\wp(\omega)$ with respect to K-standard models, and its semantic commitments are relatively modest.

3.3 *The Standard Models of Elementary Arithmetic*

It is generally acknowledged that (most) mathematical theories can be faithfully reduced to a definitional extension of set theory. If this is accepted, then the analysis of set theory developed in Sects. 2 and 3 is perfectly general. However, this is not an unproblematic thesis. For example, there are the usual, well-known arguments to the effect that set theory cannot provide an ontological reduction of mathematics.[10] Therefore, an independent account of mathematical truth for other theories is desirable. Fortunately, it is possible to provide a similar double-layer account to other foundational axiomatic theories, such as elementary arithmetic,[11] an axiomatic theory in which truth and provability also seem to be mismatched. In contrast with the theory of groups, for example, in which permutation groups are the primary phenomena, the (first-order) formal systems for set theory and elementary arithmetic cannot be obtained as axiomatizations of such classes of models that we come upon as the primary phenomena, independently of those formal systems. Instead, the formal systems for set theory and elementary arithmetic can be obtained as formal counterparts of directive principles. In the case of elementary arithmetic the double-layer schema also applies — the first layer consists of the directive principles, the second layer consists of formal systems. A list of directive principles for elementary arithmetic, formulating the notational-algorithmic conception of arithmetical operations, can be given as follows:

1. Each number is denoted by a unique numeral, which is a syntactic object obtained by a repetition, possibly null, of a primitive symbol. Each numeral denotes a unique number.
2. Given two nsumerals s and t, the sum of the numbers denoted by s and t is denoted by the numeral obtained by the repetition of the primitive symbol determined by t over s.
3. Given two numerals s and t, the product of the numbers denoted by s and t is denoted by the numeral obtained by the repetition of the repetition s determined by t.

[10]For a clear and concise exposition of this point, see [10], pp. 36–39.

[11]I am using the expression "elementary arithmetic" to designate the axiomatic theory of natural numbers without reference to sets of natural numbers. In order to designate the axiomatic theory of natural numbers and sets of natural numbers the expression "elementary analysis" is preferred.

The subject matter of elementary arithmetic can be analyzed in an analogous way: It will be proved that all structures conforming to the principles above in the obvious sense are isomorphic to the structure $(\omega, +, \times)$. Therefore, that elementary arithmetic is concerned with the standard model is, in this account, a consequence of the very directives which are the conceptual component of this axiom system, and this is very different from declaring a model standard by decree.

These directive principles prescribe a criterion for the interpretation of formal systems for elementary arithmetic. Now, there is a mathematical theorem which can be seen as evidence for the thesis that in this specific case the directive principles do what is expected, and their first-order formalizations are always *partial*:

Theorem 3 *Every structure* $\mathscr{A} = (D, \oplus, \otimes)$ *conforming to the directive principles above, in the sense that:*

1. *for every* $d \in D$ *there is a unique numeral* s *such that* $d = s^{\mathscr{A}}$,
2. *given two elements* $s^{\mathscr{A}}$ *and* $t^{\mathscr{A}}$ *in D, the sum* $s^{\mathscr{A}} \oplus t^{\mathscr{A}} = (s\hat{}t)^{\mathscr{A}}$, *where* $s\hat{}t$ *is the numeral obtained by the repetition of the primitive symbol determined by* t *over* s, *and*
3. *given two elements* $s^{\mathscr{A}}$ *and* $t^{\mathscr{A}}$ *in D, the product* $s^{\mathscr{A}} \otimes t^{\mathscr{A}} = (s * t)^{\mathscr{A}}$, *where* $s * t$ *is the numeral obtained by the repetition of the repetition* s *determined by* t*, is isomorphic to the structure* $(\omega, +, \times)$.

Proof For each $d \in D$, let s be the unique numeral such that $d = s^{\mathscr{A}}$, and let n be the unique number canonically associated with s. Clauses (2) and (3) imply that the bijection $d \mapsto n$ is an isomorphism. □

The directive principles given for elementary arithmetic can be equivalently expressed in $L_{\omega_1\omega}$. Therefore its semantic commitments are very modest since this language is arguably a slight extension of the first-order language of elementary arithmetic.

4 On Foundations of Mathematical Truth

It is usually said that set theory is a *foundation of mathematics*. I understand this as the claim: Mathematical truth can be explained away in terms of the better understood set-theoretic truth. Thus, according to this understanding of the above claim, set theory is a foundation of mathematical truth. I agree with this claim, but this is not an unproblematic thesis. In fact, as it was already mentioned, there are well-known arguments to the effect that set theory cannot provide an ontological reduction of mathematics. However, an answer to this objection is that to provide an ontological reduction is not a requirement for a foundation of mathematical truth: What is required for set theory to be a foundation of mathematical truth is that (a) all mathematical theorems can be formalized as theorems in the formal systems for set theory, and (b) this mathematical theory comes with a good understanding of its truth. Item (*a*) hardly needs to be argued for. With respect to (*b*), if the explanation of mathematical truth in terms of set-theoretic truth is to represent an improvement in clarity and precision, then set-theoretic truth must be well-understood.

In this paper I have defended an understanding of the commitments behind set-theoretic truth based on certain directive principles. These directives were subject to a mathematical analysis in order to clarify their delimitative role in the interpretation of set theory. In that mathematical analysis, the subject matter of set theory is, according to the first list of directive principles and Theorem 1, constituted by the hierarchies of sets $(V_\kappa, \in)$, in which κ is a strongly inaccessible cardinal, and according to the second list and Theorem 2, is constituted by the transitive models of ZFC which are closed under countable subsets. These were established as mathematical facts *based on Definitions* 1 and 2, respectively. Since set theory is considered a standard foundation of mathematics, these accounts of the commitments behind the interpretation of set theory correspond to accounts of the commitments behind mathematical truth. Now, it is important to finally explain how these pictures of set theory and mathematical truth solve the original problems.

The presumed concrete character of formal systems for set theory must be defined. Based on Gödel's incompleteness theorems, set-theoretic truth is generally held to transcend formal systems. If set-theoretic truth is supposed to go beyond provability in formal systems, then we have to explain how can we ground truth-values in set theory and what are we committed to when we do this. Our first task is to define what is a concrete axiom system. The proposed solution to this problem was formulated in terms of a double-layer schema of concrete axiom systems: A concrete axiom system is constituted by two layers, the layer of directive principles of the system, which correspond to the conceptual component of the system, and the layer of the formal systems, which correspond to the deductive component. Directive principles give us the commitments latent in the interpretation of formal systems, separating appropriate structures from all possible realizations of the axioms (when mathematically analyzed).

The second thing to do is to give directive principles for set theory. Two lists of directive principles were presented. The double-layer pictures of set theory and set-theoretic truth unfolded in Sects. 2 and 3 are such that truth-values are fixed beyond provability. For example, each sentence of second-order arithmetic has a rigid truth-value in each one of the double-layer pictures presented here. Of course, the truth of every theorem of 20th century classical mathematics which can be formalized in ZFC is also fixed in these double-layer pictures. That is, transferring the analysis of the commitments behind set-theoretic truth set forth in this paper to 20th century classical mathematics, it is possible to provide an appropriate account of the commitments latent in the notion of mathematical truth we happen to have inherited that does not rest on mathematical intuition, nor on formalization alone, but on principles which give us criteria for interpreting formal systems.

The proposed double-layer schema of concrete axioms systems is not committed to objects in an independent model, but it is committed to the objectivity of the criterion given by directive principles in the first layer, which corresponds to the conceptual component of the system. A given structure either conforms to the criterion or not, according to the relevant definitions of conformity, and this is what is understood by the objectivity of the criterion given by directive principles. It is a semantic commitment, it is what is required by the conceptual component of the system in order to separate some structures from all possible realizations of the axioms. The second list of directive principles given for set theory has a modest semantic commitment, when compared to the first list, but the first list seems to be closer to the concept of membership hierarchy, which is arguably the most popular candidate for the conceptual component of set theory. I agree with Kreisel's dictum[12] that the point is not the existence of objects but the objectivity of mathematical truth. Philosophy of mathematics can profit from a shift of focus from the category of the object to the notion of objectivity, and to account for set-theoretic truth in terms of an objective criterion is such a shift.

References

1. Bernays, P. 1942. *A System of Axiomatic Set Theory: Part III. Infinity and Enumerability. Analysis*. The Journal Of Symbolic Logic.
2. Cantor, G. 1899. *Letter to Dedekind*, [translated in van Heijenoort 1967, 113–117].
3. Cantor, G. 1955. *Contributions to the Founding of the Theory of Transfinite Numbers*. Dover.
4. Cohen, P. 2008. *Set Theory and the Continuum Hypothesis*. Dover.
5. Ferreirós, J. 2012. *On Arbitrary Sets and ZFC*, Bulletin of Symbolic Logic.
6. Fraenkel, A., Y. Bar-Hillel, and A. Levy. 1973. *Foundations of Set Theory*. North-Holland.
7. Gödel, K. 1947, 1963. *What is Cantor's Continuum Problem* [in Benacerraf and Putnam 1983, 470–485].
8. Gaifman, H. 2012. *On Ontology and Realism in Mathematics*. Review of Symbolic Logic.
9. Gödel, K. 1995. *Collected Works*, vol. III. Oxford.
10. McGee, V. 1997. *How Can we Learn Mathematical Language*, The Philosophical Review, vol. 106, no. 1, pp. 35–68.
11. Moore, G. 2013. *Zermelo's Axiom of Choice*. Dover.
12. Shoenfield, J. 2001. *Mathematical Logic*. ASL.
13. Tait, W. 2005. *The Provenance of Pure Reason*. Oxford.
14. Zermelo, E. 1908. *Untersuschungen über due Grundlagen der Mengenlehre, I*, Mathematische Annalen 65, 261–281 [translated in van Heijenoort 1967, 199–215].
15. Zermelo, E. 1930. Über Grenzzahlen und Mengenbereiche: Neue Untersuchungen über die Grundlagen der Mengenlehre. *Fundamenta Mathematicae* 16: 29–47.

[12]Folklore. See [8], footnote 4.

Coherence of the Product Law for Independent Continuous Events

Daniele Mundici

Abstract Let A^* and B^* be finite sets of *continuous events* (e.g., physical observables, or random variables) represented by elements of semisimple MV-algebras A and B. Suppose $\alpha\colon A^* \to [0, 1]$ and $\beta\colon B^* \to [0, 1]$ are *coherent books*, i.e., maps satisfying de Finetti's coherence criterion. Suppose all events in A^* are (logically) independent of all events in B^*. Let $C = A \otimes B$ be the semisimple tensor product of A and B. We first prove that if $a, a' \in A^*$ and $b, b' \in B^*$ satisfy $a \otimes b = a' \otimes b'$, then $\alpha(a)\beta(b) = \alpha(a')\beta(b')$. Thus by setting $\gamma(a \otimes b) = \alpha(a)\beta(b)$ we obtain a [0, 1]-valued *function* γ defined on the set C^* of pure tensors of C of the form $a \otimes b$ for $a \in A^*$ and $b \in B^*$. We then prove that γ is a coherent book on C^*. For the proofs we need the MV-algebraic extension of de Finetti Dutch Book theorem, Fubini theorem, and the Kroupa–Panti theorem (which in turn rests on the preservation properties of the Γ functor, the Stone–Weierstrass theorem and the Riesz representation theorem).

1 Main Result

We refer to [1, 15] for background on MV-algebras. The latter stand to continuously valued events (or random variables) as boolean algebras stand to yes-no events. For every MV-algebra D we let $\hom(D)$ be the set of homomorphisms of D into the standard MV-algebra [0, 1]. Let $D^* = \{d_1, \ldots, d_n\}$ be a finite subset of D. Following [9, Definition 2.1] or [15, Definition 1.1], a map $\delta\colon D^* \to [0, 1]$ is said to be $(\hom(D))$-*coherent* if

$$\text{for all } \sigma\colon D^* \to \mathbb{R} \text{ there is } \upsilon \in \hom(D) \text{ with } \sum_{i=1}^{n} \sigma(d_i)(\delta(d_i) - \upsilon(d_i)) \geq 0. \quad (1)$$

When D is clear from the context, any such map δ will be said to be a *coherent book*. In the particular case when D is a boolean algebra, (i.e., $x \oplus x = x$ for all $x \in D$),

D. Mundici (✉)
Department of Mathematics and Computer Science "Ulisse Dini",
University of Florence, Viale Morgagni 67/A, 50134 Florence, Italy
e-mail: mundici@math.unifi.it

W. Carnielli and J. Malinowski (eds.), *Contradictions, from Consistency to Inconsistency*, Trends in Logic 47, https://doi.org/10.1007/978-3-319-98797-2_10

coherent books on D^* coincide with de Finetti's coherent betting systems on the set D^* of yes-no events, in the sense of [2, pp. 311–312], [3, Chap. 1], [4, § 3.3].

Our main result in this paper shows that coherence is preserved by taking product books of coherent books on independent sets of *continuous* events. See Sect. 2 for details on the (probability-free, purely logical) notion of independence used throughout.

Theorem *Let A and B be semisimple MV-algebras and $C = A \otimes B$ their semisimple tensor product. Let A^* and B^* be finite subsets of A and B. Let*

$$C^* = \{w \in C \mid w = a \otimes b \textit{ for some } a \in A^* \textit{ and } b \in B^*\}. \tag{2}$$

Suppose $\alpha\colon A^* \to [0, 1]$ and $\beta\colon B^* \to [0, 1]$ are coherent books. Then:

(i) Whenever $a, a' \in A^*$ and $b, b' \in B^*$ satisfy $a \otimes b = a' \otimes b'$, we have the identity $\alpha(a)\beta(b) = \alpha(a')\beta(b')$.

(ii) The map

$$\gamma : a \otimes b \in C^* \mapsto \alpha(a)\beta(b) \in [0, 1], \quad (\textit{for all } a \in A^* \textit{ and } b \in B^*),$$

whose existence is ensured by (i), is a coherent book on C^*.

We say that γ is the product book *of α and β.*

Proof We let $\mu(A)$ and $\mu(B)$ denote the maximal spectral spaces of A and B. In the light of [15, § 4.5], throughout we will identify the semisimple MV-algebra A (resp., B) with a separating subalgebra of [0, 1]-valued continuous functions on $\mu(A)$, (resp., on $\mu(B)$). Further, via the representation [15, Theorem 9.17, Corollary 9.18], we also identify the maximal spectral space $\mu(C)$ with the product space $\mu(A) \times \mu(B)$. We define the *cylindrification maps*

$$\mathrm{cyl}\uparrow\colon A \to A \otimes B \ \text{ and } \ \mathrm{cyl}_{\rightarrow}\colon B \to A \otimes B$$

by stipulating that for any $f \in A$ the map cyl↑ transforms f into the function $f_{\mathrm{cyl}\uparrow}\colon \mu(A) \times \mu(B) \to [0, 1]$ given by

$$f_{\mathrm{cyl}\uparrow}(\mathfrak{m}, \mathfrak{n}) = f(\mathfrak{m}), \ \text{for all } \mathfrak{m} \in \mu(A) \text{ and } \mathfrak{n} \in \mu(B). \tag{3}$$

Similarly, for any $g \in B$,

$$g_{\mathrm{cyl}_{\rightarrow}}(\mathfrak{m}, \mathfrak{n}) = g(\mathfrak{n}), \ \text{for all } \mathfrak{m} \in \mu(A) \text{ and } \mathfrak{n} \in \mu(B). \tag{4}$$

As a main property of the semisimple tensor product, [13], [15, § 9.4], these cylindrification maps isomorphically embed A (resp., B) onto a subalgebra $A_{\mathrm{cyl}\uparrow}$ (resp., onto a subalgebra $B_{\mathrm{cyl}_{\rightarrow}}$) of C. For any $f \in A$ and $g \in B$ the product $f_{\mathrm{cyl}\uparrow} \cdot g_{\mathrm{cyl}_{\rightarrow}}$ is said to be a *pure tensor*, denoted $f \otimes g$. As proved in [15, 9.17(ii)], C is the smallest MV-algebra of (necessarily continuous) [0, 1]-valued functions on the compact Hausdorff space $\mu(C)$, containing all pure tensors.

As the reader will recall, [11, 12], for every MV-algebra D, a *state* s on D is a function $s\colon D \to [0, 1]$ satisfying $s(1) = 1$ and having the following additivity property: for all $x, y \in D$, if $x \odot y = 0$ then $s(x \oplus y) = s(x) + s(y)$. When D is a boolean algebra, a state on D is just a (finitely additive, normalized) measure on D as defined in the classical paper [5].

The ($\Rightarrow$) direction of [9, Theorem 3.2]—which is the MV-algebraic generalization of de Finetti Dutch Book theorem, [2, §§ 8–9], [3, pp. 7–8], [4, §§ 3.3, 3.4, 3.8], yields states s on A and t on B respectively extending the coherent books α on A^* and β on B^*. The Kroupa–Panti theorem [8, Corollary 29], [17, Proposition 1.1], [15, Theorem 10.5] then yields uniquely determined regular Borel probability measures λ and ν on the Borel σ-algebras $\mathscr{B}(\mu(A))$ and $\mathscr{B}(\mu(B))$ of the compact Hausdorff spaces $\mu(A)$ and $\mu(B)$ in such a way that

$$s(a) = \int_{\mu(A)} a \, \mathrm{d}\lambda \quad \text{and} \quad t(b) = \int_{\mu(B)} b \, \mathrm{d}\nu \qquad \text{for all } a \in A \text{ and } b \in B. \tag{5}$$

This theorem rests on the preservation properties of the Γ functor, [10, § 3], the Stone–Weierstrass theorem and the Riesz representation theorem. With reference to [20, § 1.7], let ξ be the product measure of the (trivially) σ-finite measures λ and ν. Combining Fubini theorem with (3)–(4), for each pure tensor $a \otimes b$ we obtain

$$\begin{aligned}\int_{\mu(C)} a \otimes b \, \mathrm{d}\xi &= \int_{\mu(C)} a_{\mathrm{cyl}\uparrow} \cdot b_{\mathrm{cyl}\rightarrow} \, \mathrm{d}\xi \\ &= \int_{\mu(B)} \left(b \int_{\mu(A)} a \, \mathrm{d}\lambda \right) \mathrm{d}\nu \\ &= \int_{\mu(A)} a \, \mathrm{d}\lambda \cdot \int_{\mu(B)} b \, \mathrm{d}\nu.\end{aligned}$$

Let the map $u\colon C \to [0, 1]$ assign to every $c \in C$ the quantity

$$u(c) = \int_{\mu(C)} c \, \mathrm{d}\xi. \tag{6}$$

A direct verification shows that u is a state of C, called *the product state* of s and t. From (5)–(6) for all $a, a' \in A$ and $b, b' \in B$ with $a \otimes b = a' \otimes b'$ we get

$$u(a \otimes b) = s(a) \cdot t(b) = s(a') \cdot t(b') = u(a' \otimes b'),$$

from which property (i) follows as a particular case. Let the map $\gamma\colon C^* \to [0, 1]$ be obtained by restricting the product state u to the set C^* of pure tensors defined in (2). By the ($\Leftarrow$)-direction of [9, Theorem 3.2], γ is a coherent book on C^*s. By construction, γ is the product book of α and β.

This settles (ii) and completes the proof of the theorem. □

In [16] the present author proved the boolean fragment of the foregoing theorem, stating that the product book of two coherent books defined on independent sets of *yes-no* events (exists and) is coherent.

2 Concluding Remarks

Suppose $A^* = \{a_1, \ldots, a_m\}$ and $B^* = \{b_1, \ldots, b_n\}$ are sets of *yes-no* events, with each a_i independent of each b_j. Formally speaking, $A^* \subseteq A$ and $B^* \subseteq B$ for *independent* subalgebras A and B of some boolean algebra C as defined in [18, § 13] or [7, Chap. 4, 11.3]. So, by definition, $0 \neq a \in A,\ \ 0 \neq b \in B$ entails $0 \neq a \wedge b \in C$. In particular, if $A \cup B$ generates C then C is canonically isomorphic to the free product $A \bigoplus B$ of A and B. Suppose $\alpha : A^* \to [0, 1]$ and $\beta \colon B^* \to [0, 1]$ are coherent books. One may naturally conjecture that the map assigning to each event (a_i and b_j) the value $\alpha(a_i)\beta(b_j)$ is a (well defined) coherent book. This conjecture, first proved in [16], also follows from our theorem in this paper, as the special case when A and B are boolean algebras, because A and B are then semisimple MV-algebras, and their (semisimple) tensor product $A \otimes B$ coincides with their free product $A \bigoplus B$, [7, § 11.1], [15, 9.17–18]. In the present boolean algebraic context, the event (a_i and b_j) amounts to the conjunction $(a_i \ \wedge\ b_j)$. Thus the mutual "independence" of the finite sets of yes-no events $A^* \subseteq A$ and $B^* \subseteq B$ (for boolean algebras A and B) may be accounted for in algebraic terms as follows: The canonical injections $A \hookrightarrow A \bigoplus B \hookleftarrow B$ isomorphically map A and B onto *independent* subalgebras A', B' of $A \bigoplus B = A \otimes B$, the latter denoting the semisimple tensor product of A and B *qua* MV-algebras.

For continuous events the situation is more delicate, because the free product of two MV-algebras is in general different from their tensor product, [15, Example 9.15, p.108]. Only the latter provides a useful formalization of "independent" sets of MV-algebraic events $A^* \subseteq A$ and $B^* \subseteq B$, via the cylindrification embeddings of A and B into $A \otimes B$. For each $a \in A^*$ and $b \in B^*$ the event (a and b) is represented by the pure tensor $a \otimes b$. The product book γ of α and β assigns to $a \otimes b$ the value $\alpha(a)\beta(b)$. The existence of γ as a *map* follows from condition (i) in our theorem, whose validity crucially depends on the assumed coherence of α and β. While Łukasiewicz logic has no connective to formalize the tensor product operation, once $A \otimes B$ is written as the Lindenbaum algebra of some theory, [15, § 1.5], as in [13, constructions, pp. 234 and 237], there do exist formulas coding each element $a \otimes b$ in the domain of γ.

De Finetti's definition (1) of a coherent book, as well as the definition of a product book, are the same for boolean and for MV-algebras. Our theorem settles the natural conjecture that coherence is preserved by taking product books on independent (continuous, as well as yes-no) events—where "independence" is given the natural probability-free definition. Needless to say, if coherence were not preserved, the notion of “stochastic independence” and/or de Finetti’s notion of “coherence” would require a radical revision.

In [15, p. 9] a large class of continuous events are given a convenient formalization in the framework of MV-algebras, generalizing the usual interpretation of two-valued events as elements of boolean algebras. As noted in [14, Remarks, pp. 240–241] and [15, p. 129], once the coherence of a book β is defined via (1) for events sitting in an MV-algebra A, coherence becomes equivalent to the extendability of β to a state s of A. The MV-algebraic counterpart of Riesz-Kakutani representation theorem [6, 19], (i.e., the Kroupa–Panti theorem [8, Corollary 29], [17, Proposition 1.1]), shows that states of A are in one-one correspondence with regular Borel probability measures on the maximal spectral space $\mu(A)$. Thus states, which are (finitely additive) functionals on MV-algebras, correspond to (countably additive) probability measures.

Letting A range over semsimple MV-algebras, $\mu(A)$ will be any possible compact Hausdorff space. Since by [9, Theorem 3.2], s is a state of A iff the restriction of s to every finite subset of A is coherent, then de Finetti’s notion of a coherent book naturally incorporates the σ-additivity axiom of Kolmogorov probability theory, and also accounts for the product law in the definition of stochastically independent events.

References

1. Cignoli, R., I.M.L. D’Ottaviano, and D. Mundici. 2000. *Algebraic Foundations of Many-Valued Reasoning, Trends in Logic*, vol. 7. Dordrecht: Kluwer Academic Publishers.
2. de Finetti, B. 1993. Sul significato soggettivo della probabilitá, Fundamenta Mathematicae, 17 (1931) 298–329. Translated into English as “On the Subjective Meaning of Probability”. In *Probabilitá e Induzione*, ed. P. Monari, and D. Cocchi, 291–321. Bologna: Clueb.
3. de Finetti, B. 1980. La prévision: ses lois logiques, ses sources subjectives, Annales de l’Institut H. Poincaré, 7 (1937) 1–68. Translated into English by Henry E. Kyburg Jr., as “Foresight: Its Logical Laws, its Subjective Sources”. In *Studies in Subjective Probability, Wiley, New York, 1964*, ed. H.E. Kyburg Jr., and H.E. Smokler, 53–118. New York: Second edition published by Krieger.
4. de Finetti, B. 2017. *Theory of Probability, A Critical Introductory Treatment, Translated by Antonio Machí and Adrian Smith*. Chichester, UK: Wiley.
5. Horn, A., and A. Tarski. 1948. Measures in Boolean algebras. *Transactions of the American Mathematical Society* 64: 467–497.
6. Kakutani, S. 1941. Concrete representation of abstract (M)-spaces, (A characterization of the space of continuous functions). *Annals of Mathematics*, (2), 42: 994–1024.
7. Koppelberg, S. 1989. General Theory of Boolean Algebras. In *Handbook of Boolean Algebras* (Vol.1), ed. by J.D.Monk with the cooperation of R. Bonnet, Elsevier, Amsterdam.
8. Kroupa, T. 2006. Every state on semisimple MV-algebra is integral. *Fuzzy Sets and Systems* 157: 2771–2782.
9. Kühr, J., and D. Mundici. 2007. De Finetti theorem and Borel states in [0,1]-valued algebraic logic. *International Journal of Approximate Reasoning* 46: 605–616.
10. Mundici, D. 1986. Interpretation of AF C^*-algebras in Łukasiewicz sentential calculus. *Journal of Functional Analysis* 65: 15–63.

11. Mundici, D. 1993. Logic of infinite quantum systems. *International Journal of Theoretical Physics* 32: 1941–1955.
12. Mundici, D. 1995. Averaging the truth value in Łukasiewicz sentential logic. *Studia Logica, Special issue in honor of Helena Rasiowa* 55: 113–127.
13. Mundici, D. 1999. Tensor products and the Loomis-Sikorski theorem for MV-algebras. *Advances in Applied Mathematics* 22: 227–248.
14. Mundici, D. 2009. Interpretation of de Finetti coherence criterion in Łukasiewicz logic. *Annals of Pure and Applied Logic* 161: 235–245.
15. Mundici, D. 2011. *Advanced Łukasiewicz calculus and MV-algebras, Trends in Logic*, vol. 35. Berlin: Springer.
16. Mundici, D. 2017. Coherence of de Finetti coherence. *Synthese* 194: 4055–4063.
17. Panti, G. 2009. Invariant measures in free MV-algebras. *Communications in Algebra* 36: 2849–2861.
18. Sikorski, R. 1960. *Boolean Algebras*, 2nd edition, Ergeb. der Math. und ihrer Grenzgeb., vol. 25. Berlin: Springer.
19. Tao, T. 2010. *An Epsilon of Room I: Real Analysis*, vol. 117, Graduate Studies in Mathematics. Providence RI: American Mathematical Society.
20. Tao, T. 2011. *An Introduction to Measure Theory*, vol. 126, Graduate Studies in Mathematics. Providence RI: American Mathematical Society.

A Local-Global Principle for the Real Continuum

Olivier Rioul and José Carlos Magossi

Abstract We discuss the implications of a *local-global* (or *global-limit*) principle for proving the basic theorems of real analysis. The aim is to improve the set of available tools in real analysis, where the local-global principle is used as a unifying principle from which the other completeness axioms and several classical theorems are proved in a fairly direct way. As a consequence, the study of the local-global concept can help establish better pedagogical approaches for teaching classical analysis.

1 Introduction

The logical foundations of mathematical analysis were developed at the end of 19th century and beginning of 20th century by mathematicians such as B. Bolzano (∼1817), A. L. Cauchy (∼1821–1829), K. Weierstrass (∼1865-1895), C. Méray (∼1869), R. Dedekind (∼1872), G. Cantor (∼1872), E. Heine (∼1872), E. Borel (∼1895-1903), P. Cousin (∼1895) and H. Lebesgue (∼1905). They departed from the geometric intuition of the "real line" by establishing rigorous proofs based on *completeness axioms* that characterize the real number continuum.

As noticed in [4, 14, 27], rigor was not the most pressing question. Instead these authors focused on *teaching*. Several mathematicians found themselves in an awkward situation when they had to teach differential and integral calculus based on fuzzy geometric evidences. Therefore, they decided to reform it [27]. Examples are Cauchy's Cours d'Analyse at École Polytechnique in Paris, Weierstrass's lectures

O. Rioul (✉)
Télécom ParisTech, Université Paris-Saclay, Paris, France
e-mail: rioul@enst.fr

J. C. Magossi
Faculdade de Tecnologia, Unicamp, Limeira, Brasil
e-mail: magossi@ft.unicamp.br

W. Carnielli and J. Malinowski (eds.), *Contradictions, from Consistency to Inconsistency*, Trends in Logic 47, https://doi.org/10.1007/978-3-319-98797-2_11

at the University of Berlin and Dedekind's course at Zürich Polytechnic. Dedekind wrote:

> *In discussing the notion of the approach of a variable magnitude to a fixed limiting value, and especially in proving the theorem that every magnitude which grows continually, but not beyond all limits, must certainly approach a limiting value, I had recourse to geometric evidences.(...) For myself this feeling of dissatisfaction (...) I should find a purely arithmetic and perfectly rigorous foundation for the principles of infinitesimal analysis.* [8, pp. 1–2]

Felix Klein coined the phrase "the arithmetizing of mathematics" [22], a classic of the era of rigor. Until today, the foundation has not been put into question; it is still recognized as satisfactory in all classical textbooks which define $\mathbb{R}$ as any ordered field satisfying one of the equivalent *completeness axioms* listed below[1]:

Sup (Least Upper Bound Property) Any set of real numbers has a supremum (and an infimum)[2];
Cut (Dedekind's Completeness) Any cut defines a (unique) real number;
Nest+Arch (Cantor's Property) Any sequence of nested closed intervals has a common point + Archimedean property;
Cauchy+Arch (Cauchy's Completeness) Any Cauchy sequence converges + Archimedean property;
Mono (Monotone Convergence) Any monotonic sequence has a limit[2];
BW (Bolzano–Weierstrass) Any infinite set of real numbers (or any sequence) has a limit point[2];
BL (Borel–Lebesgue) Any cover of a closed interval by open intervals has a finite subcover[3];
Cousin (Cousin's partition [13]) Any gauge defined on a closed interval admits a fine tagged partition of this interval;
Ind (Continuous Induction [5, 17, 20]).

One may find it striking that all these equivalent properties look so diverse. This calls for the need of a simple unifying principle from which all such properties could be easily and directly derived as theorems. In this article, we introduce and discuss two versions of yet another equivalent axiom:

LG (Local-Global) Any *local* and *additive* property is *global*;
GL (Global-Limit) Any *global* and *subtractive* property has a *limit point*.

The earliest reference we could find that explicitly describes this principle is Guyou's little-known French textbook [16]. Guyou wrote:

> *Les démonstrations de ce livre sont, en général, différentes des démonstrations classiques; un tel remaniement comporte sans doute des erreurs, que je serai reconnaissant à mes collègues de bien vouloir me signaler.* [16, p. xv].

---

[1] Precise definitions will be given in Sect. 4. Some of the statements require the Archimedean property: Any real number is upper bounded by a natural number.

[2] Possibly infinite, e.g., $\sup\ \mathbb{R} = \inf\ \varnothing = +\infty$, $\sup\ \varnothing = \inf\ \mathbb{R} = -\infty$, $\lim\ \pm n = \pm\infty$.

[3] This is Borel's statement, also (somewhat wrongly) attributed to Heine, and later generalized by Lebesgue and others [1].

[The proofs in this book are, in general, different from the classical proofs; such a reworking may contain errors, that I shall be grateful to my colleagues for pointing out to me.]

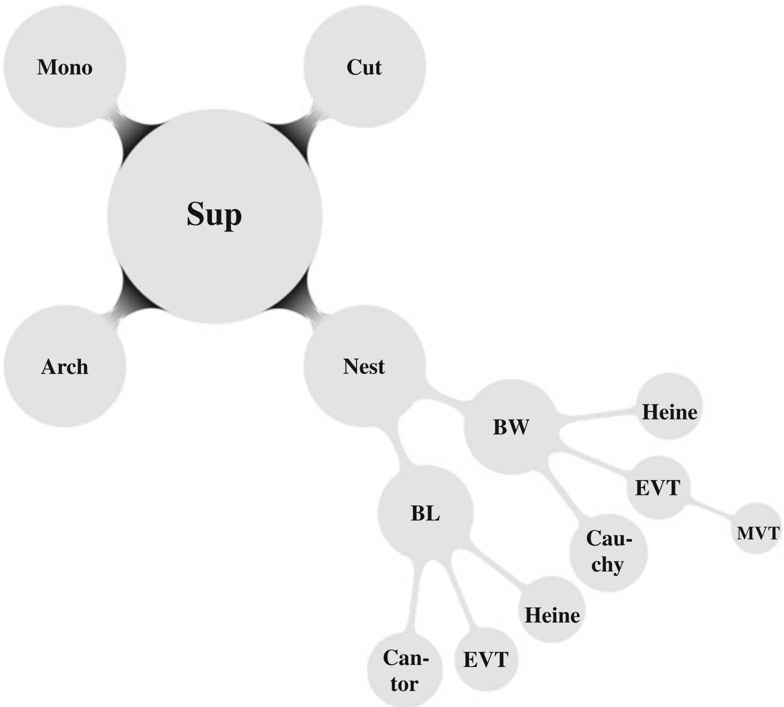
The principle was later re-discovered independently and frequently in many disguises in some American circles [11, 21, 26, 28, 30]. One central concept is the notion of interval-additive property set up independently by Guyou and Ford [11, 16], which we feel can be useful for pedagogical purposes:

A statement P concerning intervals will be called interval-additive if whenever P is true for each of two overlapping intervals [...] *it is also true for the interval obtained by combining them; that is, their union.* [11, p. 106]

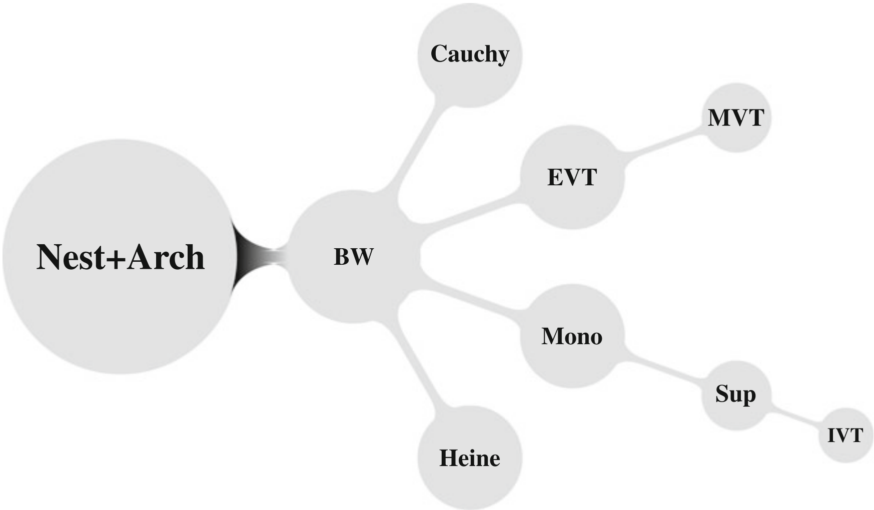
2 Teaching Real Analysis: Present Situation

We have studied the logical flow of proofs in detail in the most influential undergraduate/graduate textbooks in the U.S.A. [2, 7, 29], France [9, 25] and Brazil [15, 24]. These included not only proofs of the essential properties of the real numbers, but also of the basic theorems for continuity (boundness theorem **BT**, intermediate value theorem **IVT**, extreme value theorem **EVT**, Heine's uniform continuity theorem **Heine**) and differentiation (essentially the mean value theorem **MVT**).

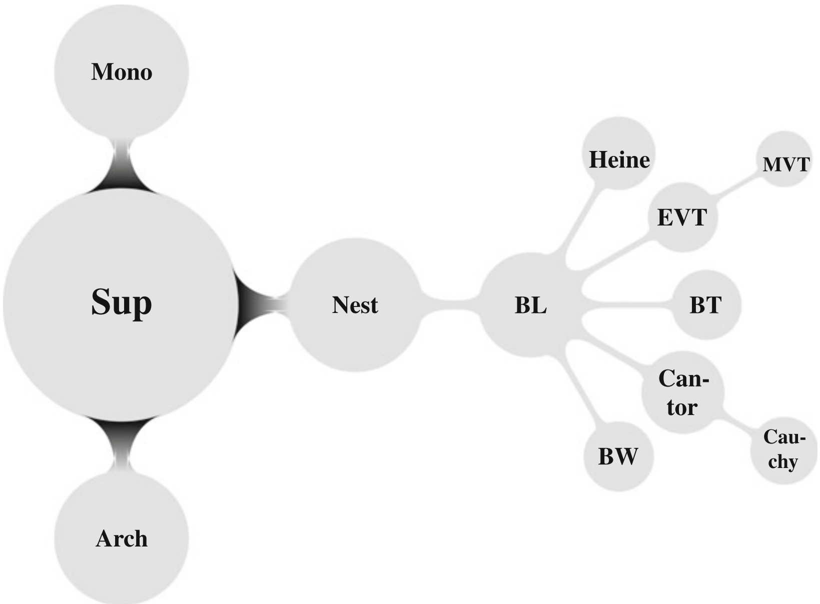
We identified the logical flows of each of these textbooks as follows:
Robert G. Bartle, Elements of Real Analysis [2]



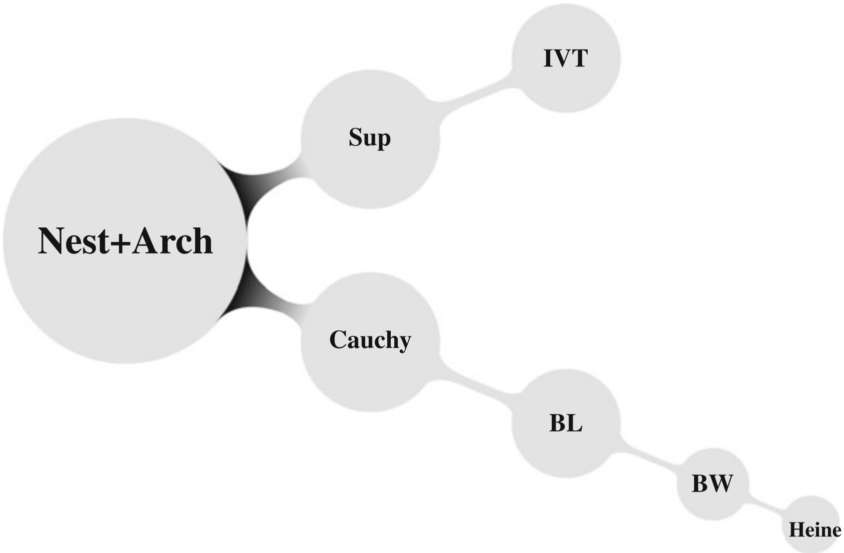
Richard Courant, Introduction to Calculus and Analysis I [7]



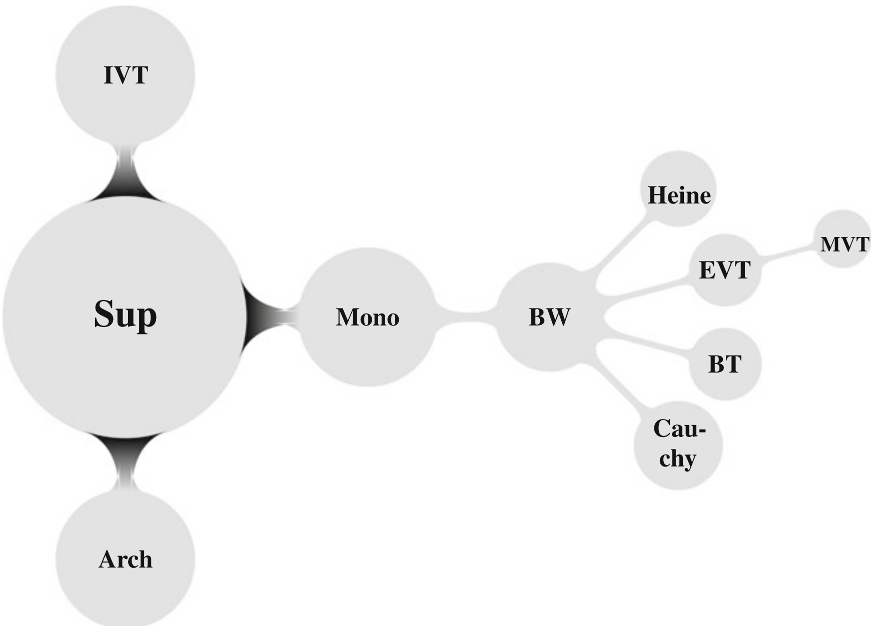
Walter Rudin, Principles of Mathematical Analysis [29]



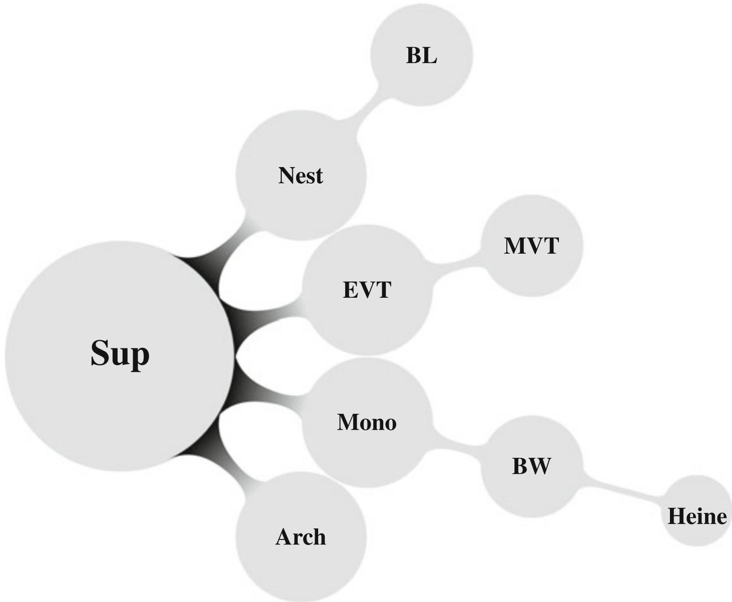
Jean Dieudonné, Foundations of Modern Analysis [9]



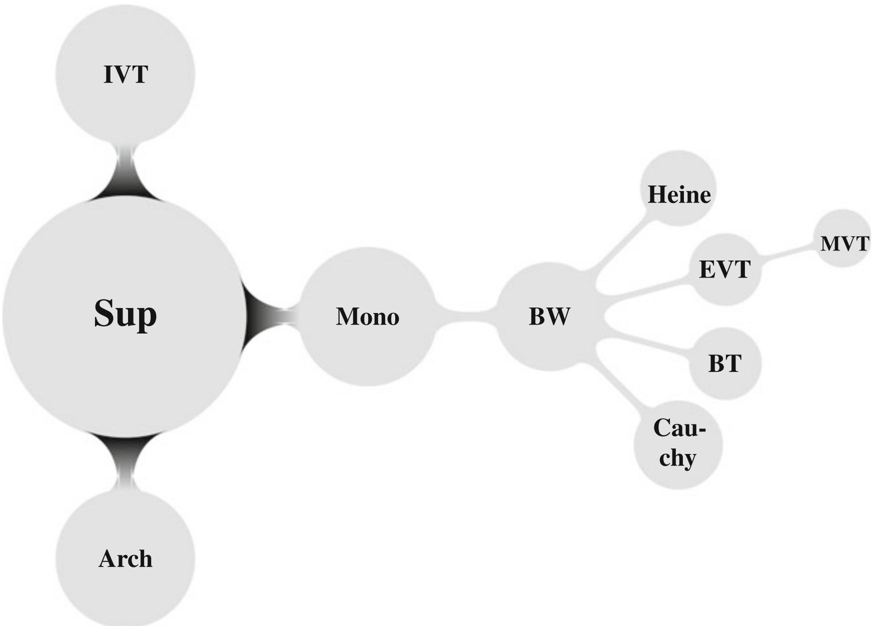
Serge Lang, Undergraduate Analysis [25]



Elon Lages Lima - Análise Real [24]



Djairo Guedes de Figueiredo - Análise I [15]

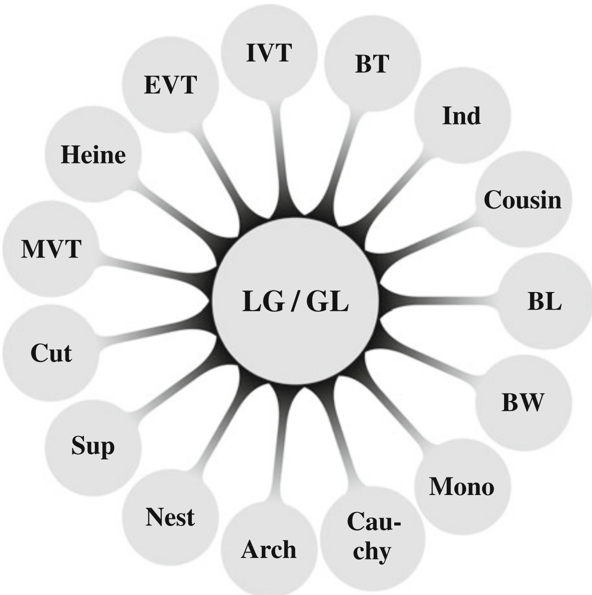


It appears that **Sup** is by far the preferred axiom, with **Nest+Arch** being the only considered alternative in [7, 9]. Other axioms (**Cauchy, Arch**, **Mono**, **BW**, often **BL**, and sometimes **Cut**) are derived as theorems. In contrast, **Cousin** and **Ind** are never used.[4] **BW** is often central to prove the basic theorems of real analysis (particularly **IVT**, **EVT**, **Heine**) with **BL** sometimes used as a "topological" alternative.

Our conclusion about classical approaches is that several classical proofs are quite difficult and subtle for the beginner (e.g., proofs of **EVT** or **Heine** using **BW**). Recent attempts to improve this situation in the literature advocate the use of **Cousin** [13] or **Ind** [5, 17], but this can also be cumbersome at times (although we agree that this is a matter of opinion).

Some textbooks (such as [29]) also mention the possibility of *proving* the fundamental axiom by first *constructing* the reals from the rationals—themselves constructed from the natural numbers— the two most popular construction methods being Dedekind's cuts and Cantor's fundamental sequences. While this is satisfactory for logical consistency, the details are always tedious and not very instructive for the student or for anyone using the real numbers, since the way they can be constructed never influences the way they are actually used.

In the following sections we describe the **LG/GL** alternative, which we show is one basic unifying principle from which all other completeness axioms and basic continuity and differentiability theorems are easily derived, as illustrated in the following figure. In this way a teacher may advantageously choose to teach (or not to teach) any given item.



[4] Although proposed at the same time as Borel's **BL**, Cousin has been largely overlooked since. It was only recently re-exhumed as a fundamental lemma for deriving the gauge (Kurzweil–Henstock) integral (e.g. [13]). **Ind** is much more recent and in fact inspired from **LG** (see [10, 19]).

3 The Local-Global Principle: A Primer

In the remainder of this paper (perhaps with the exception of Sect. 5), our presentation is deliberately at the simplest undergraduate level. In particular, we do not use explicitly topological concepts such as compactness and connectedness (even though these could be easily addressed within the present framework) and start with what we think of as the simplest type of point sets, namely intervals. We also stay one-dimensional although the concepts derived here can be easily generalized to point sets in any dimension.

Let us explain the above **LG** and **GL** principles by defining the following intuitive notions.

3.1 Points and Intervals

We follow the classical notations of points and intervals with some non-traditional definitions which will now be explained. We add two new symbols $-\infty$ and $+\infty$ to the usual real set $\mathbb{R}$ such that:

$$\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, +\infty\}$$

is totally ordered where $-\infty$ and $+\infty$ are the smallest and largest elements of $\mathbb{R}$, respectively.

Definition 1 A real *point* is an element of $\overline{\mathbb{R}}$. A finite point is an element of $\mathbb{R}$, and an infinite point is $-\infty$ or $+\infty$.

We feel that introducing explicitly *infinite* points is quite convenient here because it allows for simpler statements where e.g., we can dispense with assuming that a given point set is *bounded* to prove the existence of some limit point (since that limit point can be infinite). In other words, our take on the classical debate of potential vs. actual infinite leans toward adopting actual infinities, essentially for convenience.

Most of the sets considered in the sequel will be point sets in $\overline{\mathbb{R}}$, particularly intervals. We consider two kinds of intervals:

Definition 2 Let u, v points in $\overline{\mathbb{R}}$ such that $u < v$. The following are the two kinds of intervals with u and v as extremities:

$$[u, v] = \{x \in \overline{\mathbb{R}} | u \leq x \leq v\}$$
$$]u, v[= \{x \in \overline{\mathbb{R}} | u < x < v\}$$

To simplify the assertions we consider any $[a, b] \subseteq [-\infty, +\infty] = \overline{\mathbb{R}}$ and assume that all closed intervals $[u, v] \subseteq [a, b]$ are nondegenerate ($u < v$). Again this implicit convention of nondegenerate intervals will appear quite convenient in what follows.

Definition 3 Intervals are *overlapping* if their intersection is an interval.[5]

Central in our study are properties of intervals $[u, v] \subseteq [a, b]$. Guyou wrote:

> *Une fonction $f(x)$, qui possède la propriété d'être bornée (nous appellerons cette propriété P) dans deux intervalles contigus, possède la même propriété dans l'intervalle somme des deux (nous dirons que P est additive).* [16, p. 32]
>
> *[A function $f(x)$, which has the property of being bounded (we shall call this property P) in two contiguous intervals, has the same property in the sum of the two intervals (we shall say that P is additive).]*

Here we shall always consider properties $\mathscr{P}$ of such intervals, and any property $\mathscr{P}$ is identified with the set of intervals $[u, v]$ that satisfies this property. Thus we write "$[u, v] \in \mathscr{P}$" if $[u, v]$ satisfies the property $\mathscr{P}$. The negation $\neg\mathscr{P}$ of property is identified to the complementary set of $\mathscr{P}$, that is, $[u, v] \notin \mathscr{P} \iff [u, v] \in \neg\mathscr{P}$.

Since we identify $\mathscr{P}$ with a family of closed subintervals of $[a, b]$, the set of all properties $\mathscr{P}$ can be thought as the set-theoretic abstraction of the set of all such families of closed subintervals. It is of course much simpler to think of the statement $[u, v] \in \mathscr{P}$ as a property satisfied by an interval $[u, v]$ which is itself characterized by two endpoints $u < v$. Thus $[u, v] \in \mathscr{P}$ can be simply thought as a binary relation $u\mathscr{R}v$ on the set of ordered endpoints $u < v$.

3.2 Additive and Subtractive Properties

Definition 4 A property $\mathscr{P}$ is *additive* if for any $u < v < w$

$$[u, v] \in \mathscr{P} \ \wedge \ [v, w] \in \mathscr{P} \implies [u, w] \in \mathscr{P}.$$

A consequence of $\mathscr{P}$ being additive is that it defines a transitive binary relation $u\mathscr{R}v \iff [u, v] \in \mathscr{P}$.

Definition 5 A property $\mathscr{P}$ is *subtractive* if for any $u < v < w$,

$$[u, w] \in \mathscr{P} \implies [u, v] \in \mathscr{P} \ \vee \ [v, w] \in \mathscr{P}.$$

Proposition 1 *A property $\mathscr{P}$ is* additive *if and only if its negation $\neg\mathscr{P}$ is subtractive.*

Proof The contraposition of statement $[u, v] \in \mathscr{P} \wedge [v, w] \in \mathscr{P} \implies [u, w] \in \mathscr{P}$ is $[u, w] \in \neg\mathscr{P} \implies [u, v] \in \neg\mathscr{P} \vee [v, w] \in \neg\mathscr{P}$. □

Example 1 The property of a function being positive (or nondecreasing, or continuous) on $[u, v]$ is additive and subtractive (and in fact true for any subinterval).

[5] A non-degenerate interval. Hence two adjacent intervals $[u, v]$ and $[v, w]$ (where $u < v < w$) are not overlapping since their intersection is reduced to a point.

Example 2 The property that $[u, v]$ has exactly one integer is subtractive but not additive. The property that $[u, v]$ has at least two integers is additive but not subtractive.

Example 3 An important example of a quite general subtractive property is "$u \notin E$ and $v \in E$" where E is any set of points. In fact, if $u < v < w$, and if $[u, w] \in \mathscr{P}$, we have $[u, v] \in \mathscr{P}$ or $[v, w] \in \mathscr{P}$ according to $v \in E$ or not. As an example, for all real function f, the property "$f(u) \leq 0$ and $f(v) > 0$" is subtractive. Similarly, the property "$f(u) \leq 0$ and $f(v) \geq 0$" is also subtractive.

Certain properties are true under slightly different conditions that require a partial overlap between the two subintervals:

Definition 6 A property is *o-additive* (overlap-additive) if for all $t < u < v < w$,

$$[t, v] \in \mathscr{P} \wedge [u, w] \in \mathscr{P} \implies [t, w] \in \mathscr{P}.$$

Example 4 The property "$(v - u) \geq 1$" is additive and o-additive. The property "$(v - u) \leq 1$" is neither additive, nor o-additive.

Example 5 The property "f is a linear function on $[u, v]$" is o-additive, but not additive (consider a piecewise linear function). The property "f is a convex function on $[u, v]$" is o-additive, but not additive.

Example 6 The property "$f(u)f(v) > 0$" ("$f(u)$ and $f(v)$ are of the same sign") is additive, but not o-additive.

Definition 7 A property $\mathscr{P}$ is *o-subtractive* if for all $t < u < v < w$,

$$[t, w] \in \mathscr{P} \implies [t, v] \in \mathscr{P} \vee [u, w] \in \mathscr{P}.$$

Example 7 Any property such that, when true for an interval, remains true for any subinterval ($[u', v'] \subset [u, v] \in \mathscr{P}$ implies $[u', v'] \in \mathscr{P}$) is *a fortiori* o-subtractive.

Example 8 The property "$(v - u) < 1$" is subtractive and o-subtractive. The property "$(v - u) > 1$" is neither subtractive nor o-subtractive.

Proposition 2 *A property $\mathscr{P}$ is o-additive if and only if its negation $\neg\mathscr{P}$ is o-subtractive.*

Proof The contraposition of "$[t, v] \in \mathscr{P} \wedge [u, w] \in \mathscr{P} \implies [t, w] \in \mathscr{P}$" is "$[t, w] \in \neg\mathscr{P} \implies [t, v] \in \neg\mathscr{P} \vee [u, w] \in \neg\mathscr{P}$". □

Example 9 The property "f is a nonlinear function on $[u, v]$" is o-subtractive but not subtractive.

Example 10 The property $f(u)f(v) \leq 0$ ("$f(u)$ and $f(v)$ have opposite signs") is subtractive but not o-subtractive.

3.3 Neighborhoods

Instead of using the general notion of a neighborhood in an abstract topological space we use the following equivalent notion for $\overline{\mathbb{R}}$, based on intervals, which is enough for our purposes. A neighborhood $V(x)$ of a point $x \in [a, b]$ contains all points "sufficiently close" to x:

Definition 8 A *neighborhood* $V(x)$ of a point $x \in [a, b]$ is any set of points containing at least one interval $[u, v]$ such that:

$$\begin{cases} u < x < v, & \text{if} \quad a < x < b; \\ a = u < x, & \text{if} \quad x = a; \\ u < v = b, & \text{if} \quad x = b. \end{cases}$$

Thus in a neighborhood $V(x)$, it is possible to approach x from both sides if x belongs to the interior of $[a, b]$, but only from one side if x is one of the extremities of $[a, b]$. The notion of neighborhood depends on the considered set $[a, b]$. In most situations one may consider only neighborhoods that are themselves intervals.

Definition 9 An interval $[u, v]$ is *adapted* to neighborhood $V(x)$ if $x \in [u, v] \subseteq V(x)$.

3.4 Local Properties and Limit Points

Definition 10 A property $\mathscr{P}$ is *local* at x if there exists a neighborhood $V(x)$ such that all intervals adapted to x satisfy $\mathscr{P}$, i.e.,

$$\exists V(x), \forall [u, v] \text{ adapted to } V(x), [u, v] \in \mathscr{P}.$$

A property $\mathscr{P}$ is *local* on a set of points E if it is local at any point in E.

Example 11 As will be seen later, continuity and differentiability of functions are local properties. For example, a function f is continuous iff for any $\varepsilon > 0$, "$|f(u) - f(v)| < \varepsilon$" is local. Some topological properties such as interior point or isolated point can also be seen as local properties.

Definition 11 A property $\mathscr{P}$ has a *limit* point x if each neighborhood $V(x)$ contains an adapted interval which satisfies $\mathscr{P}$, i.e.,

$$\forall V(x), \exists [u, v] \text{ adapted to } V(x), [u, v] \in \mathscr{P}.$$

A property $\mathscr{P}$ has a *limit* on a set of points E if it has a limit at each point of E. It can be easily seen as an exercise that any property local at x does have a limit at x.

Proposition 3 *A property $\mathscr{P}$ is not* local *at x if and only if its negation $\neg\mathscr{P}$ has a limit at x.*

Proof The negation of the assertion $\exists V(x), \forall[u, v]$ adapted to $V(x)$, $[u, v] \in \mathscr{P}$ is $\forall V(x), \exists[u, v]$ adapted to $V(x)$, $[u, v] \in \neg\mathscr{P}$. □

Thus to say that $\mathscr{P}$ is *not* local on a set E is the same as saying that $\neg\mathscr{P}$ has a limit at *at least* one point of E.

3.5 Local-Global and Global-Limit Axioms

We introduce the following as foundations (completeness axioms) for the real numbers. As shown later in Sects. 4 and 5 (Proposition 9), any one of these axioms is enough to characterize $\mathbb{R}$ or $\overline{\mathbb{R}}$ just as it is done traditionally by other completeness axioms.[6] Again let $[a, b]$ be any interval in $\overline{\mathbb{R}}$.

Local-Global Axiom (LG). *Every* local *and* additive *property on $[a, b]$ is global, that is, satisfied by $[a, b]$.*
Global-Limit Axiom (GL). *Every* global *and* subtractive *property has a* limit *point in $[a, b]$.*

Proposition 4 *The LG and GL axioms are equivalent.*

Proof Let $\mathscr{P}$ be additive, that is, $\neg\mathscr{P}$ is subtractive. The LG axiom can then be written as: if $\mathscr{P}$ is local in $[a, b]$ then $[a, b] \in \mathscr{P}$. This is in turn equivalent by contraposition to the statement: if $[a, b] \in \neg\mathscr{P}$ then $\neg\mathscr{P}$ has a limit point in $[a, b]$, which is the GL axiom. □

Lemma 1 *Any property that is both local and o-additive is additive.*

Proof Suppose $[u, v] \in \mathscr{P}$ and $[v, w] \in \mathscr{P}$. Since $\mathscr{P}$ is local at v, there exists a neighborhood of v in which any interval $[r, s]$ which contains v satisfies $\mathscr{P}$. We may then assume that $u < r < v < s < w$. Since $\mathscr{P}$ is o-additive, $[u, v], [r, s] \in \mathscr{P}$ implies $[u, s] \in \mathscr{P}$, then $[u, s], [v, w] \in \mathscr{P}$ implies $[u, w] \in \mathscr{P}$. Thus $\mathscr{P}$ is additive. □

It follows from the lemma that in the LG axiom, we may always consider either additive or o-additive properties. Thus we obtain the equivalent variants:

Local-Global Axiom (LG)–variant. *Every* local *and o-*additive *property is* global, *that is, satisfied by $[a, b]$.*
Global-Limit Axiom (GL)–variant. *Every* global *and o-*subtractive *property has a* limit *point in $[a, b]$.*

[6]Even though the proposed axioms appear to be second-order statements since they are quantified over properties of sets, in fact any property is simply identified to a family of subintervals. Therefore, the axioms only require the basic (first order) ZF theory (with or without the axiom of choice), as is usual when teaching real analysis at an elementary level.

4 Elementary Theorems for the Reals

Many common mathematical properties can be identified as local and many common mathematical objects can be identified as limit points. For example, a function f is continuous iff for any $\varepsilon > 0$, "$|f(u) - f(v)| < \varepsilon$" is local; a sequence x_k converges iff "$x_k \in [u, v]$ for sufficiently large k" has a limit point. Thus, taking LG or GL as the fundamental completeness axiom for the real numbers it becomes easy and intuitive to prove all the other completeness properties, as well as all the basic theorems of real analysis. We start with the elementary theorems for the reals.

4.1 Dedekind Cuts

Definition 12 A *Dedekind cut* is a pair (E, E') where point set E and its complement set E' in $[a, b]$ are such that $E' < E$, that is, $u < v$ for any $v \in E$ and $u \notin E$.

Theorem 1 (**Cut**: Dedekind's Completeness) *Any cut* (E, E') *defines a (unique) point* x *such that* $E' \leq x \leq E$.

Proof We can assume by hypothesis that E and E' are non-empty sets. The property $[u, v] \in \mathscr{P}$ with $u \notin E$ and $v \in E$ is global and also subtractive (see Example 3). By the GL axiom, $\mathscr{P}$ has a limit point x: any neighborhood $V(x)$ contains $u < x < v$ such that $u \notin E$ and $v \in E$. From this we can deduce that no point $x' \in E$ is $< x$, otherwise we could find $u \in E'$ so that $x' < u < x$, which contradicts the hypothesis $E' < E$. Similarly no point in E' is $> x$. Hence $E' \leq x \leq E$. □

For completeness we observe the following.

Proposition 5 *The LG axiom is equivalent to Dedekind's completeness.*

Proof It is enough to prove the GL axiom from Dedekind's completeness theorem. Let $\mathscr{P}$ be global and subtractive and E be the set of points v for which $[a, v'] \in \mathscr{P}$ for all $v' \geq v$. Clearly $a \notin E$ (since $[a, a]$ is not a non-degenerate interval) and $b \in E$. Since $v \in E$ implies that all $v' \geq v$ are in E, one has $E' \leq E$ and (E, E') is a cut, and there exists x such that $E' \leq x \leq E$. In every neighborhood $V(x)$ one can find $[u, v]$ such that $[a, v] \in \mathscr{P}$ but $[a, u] \notin \mathscr{P}$. Since $\mathscr{P}$ is subtractive, $[u, v] \in \mathscr{P}$. Hence $\mathscr{P}$ has limit point x. □

4.2 Supremum and Infimum

Instead of the usual definitions of supremum (least upper bound) and infimum (greatest lower bound), we may use the following definitions which are easily shown to be equivalent and are more convenient for our purposes.

Definition 13 (*Supremum*) A point set E has *upper bound* v if $E \leq v$. It has *supremum* $x = \sup E$ if the property that "v is an upper bound of E and u is not an upper bound of E" has limit point x.

Lower bound and infimum are defined similarly.

Theorem 2 (**Sup**: Least Upper Bound Property) *Every set of points has a supremum.*

Proof We may assume that the set E is non empty (otherwise $\sup E = -\infty$) and not reduced to $\{a\}$ (in which case $\sup E = a$). The above-mentioned property $\mathscr{P}$: "$v \geq E, u \ngeq E$" is global since $b \geq E$ and $a \ngeq E$. It is also clearly subtractive (see Example 3). By the GL axiom, $\mathscr{P}$ has a limit point which is $\sup E$. □

For completeness we observe the following.

Proposition 6 *The LG axiom is equivalent to the supremum theorem.*

Proof It is enough to prove the LG axiom from the supremum theorem. Let $\mathscr{P}$ be local and additive and set $s = \sup\{x \mid [a, x] \in \mathscr{P}\}$. Since $\mathscr{P}$ is local in s, each interval $[u, v]$ adapted to a neighborhood $V(s)$ satisfies $\mathscr{P}$. If $s < b$, one can choose $[u, v]$ such that $v > s$ and $[a, u] \in \mathscr{P}$. Then by additivity, $[a, v] \in \mathscr{P}$ which contradicts the definition of s. Therefore $s = b$ and $[a, b] \in \mathscr{P}$. □

4.3 Continuous Induction Principle

In this section we consider properties on *points* in $[a, b]$. Such a property P can be identified with a set of points satisfying this property; then $x \in P$ is just a short notation for "x satisfies P". The principle of mathematical induction in $\mathbb{N}$ concerns the properties of natural numbers. The property $P \subset \mathbb{N}$ is *inductive* if:

1. $0 \in P$;
2. if $n > 0$ and if all $k < n$ satisfies P, then $n \in P$;
3. if $n \in P$, then $n + 1 \in P$.

Due to the discrete nature of the integers, condition 2 implies condition 3, so we may only assume conditions 1 and 3 (the usual induction) or conditions 1 and 2 (the "strong induction"). The principle of mathematical induction says that any inductive property is satisfied for all integers: P inductive $\Longrightarrow$ $P = \mathbb{N}$. This principle can now be stated for the real numbers as follows:

Definition 14 A property P (subset of $[a, b]$) is *inductive* if

1. $a \in P$;
2. if $x > a$, and if all $u < x$ satisfies P, then $x \in P$;
3. if $x < b$, and if $x \in P$, then there exists $v > x$ such that $[x, v] \subset P$.

Theorem 3 (**Ind**: Continuous Induction (or Real Induction [5])) *Any inductive property P is global. (In other words, the only induction subset of* $[a, b]$ *is* $[a, b]$ *itself.)*

Proof Let $\mathscr{P}$ be the property of intervals $[u, v]$ defined by "each point $\leq u$ satisfies P but there exists a point $\leq v$ which does not satisfy P". This property is subtractive (see Example 3). Since $a \in P$, if it is not true that any point $\leq b$ satisfies P, then $\mathscr{P}$ would be global. Assume by contradiction that this is the case. By the GL axiom, $\mathscr{P}$ admit a limit point x. Then all $u < x$ satisfies P and by condition 2, $x \in P$. If $x < b$, by condition 3, we find $v > x$ such that each point $\leq v$ satisfies P, which contradicts x as a limit point of $\mathscr{P}$. Therefore the limit point is equal to b, and each point $\leq b$ satisfies P. □

For completeness we observe the following.

Proposition 7 *The LG axiom is equivalent to the continuous induction principle.*

Proof It is enough to prove the LG axiom from the continuous induction principle. Let $\mathscr{P}$ be local and additive and let P be the set of x points for which either $x = a$ or $[a, x] \in \mathscr{P}$. By hypothesis $a \in P$.

- Let $x > a$ and assume that all points $< x$ satisfy P. Since $\mathscr{P}$ is local in x, each interval $[u, v]$ adapted to a certain neighborhood $V(x)$ satisfies $\mathscr{P}$. Since $x > a$ we can choose $[u, v]$ such that $a < u < x$ and $v = x$. Then $u \in P$. Both intervals $[a, u]$, $[u, x]$ satisfy $\mathscr{P}$. Since $\mathscr{P}$ is additive, $[a, x]$ satisfies $\mathscr{P}$, so $x \in P$.
- Let $x < b$ and assume that $x \in P$. As $\mathscr{P}$ is local in x, each interval $[u, v]$ adapted to a certain neighborhood $V(x)$ satisfies $\mathscr{P}$. Since $x < b$ we can choose $[u, v]$ so that $u = x$ and $v > x$. For all v' such that $x < v' \leq v$, $[x, v']$ remains adapted to $V(x)$, and so satisfies $\mathscr{P}$. If $a = x$, we have $[a, v'] \in \mathscr{P}$. Otherwise $a < x$, and the two intervals $[a, x]$, $[x, v']$ satisfy $\mathscr{P}$. Since $\mathscr{P}$ is additive, $[a, v'] \in \mathscr{P}$. In both cases $[a, v'] \in \mathscr{P}$ for all v' such that $x < v' \leq v$, hence $[x, v] \subset P$.

By continuous induction we deduce that $b \in P$, that is, $[a, b] \in \mathscr{P}$. □

4.4 Monotone Limits

Instead of the usual definition of a limit of a sequence we can adopt the following definition which is easily shown to be equivalent and is more convenient for our purposes.

Definition 15 A sequence (x_k) of points has *limit* ℓ: $x_k \to \ell$ if the property $\mathscr{P}$ that "$[u, v]$ contains all x_k for large enough k" has limit point ℓ.

Notice that ℓ can be either finite or infinite.

Theorem 4 (**Mono**: Monotone Convergence) *Any monotonic sequence has a limit.*

Proof Let (x_k) be a monotonic sequence of points in $[a, b]$. We may assume that (x_k) is nondecreasing (otherwise consider $(-x_k)$). The above-mentioned property "$[u, v]$ contains all x_k are large enough k" is clearly global and amounts to say that "v is greater than or equal to all x_k and u is not greater than or equal to all x_k". This property is subtractive (see Example 3). By the GL axiom, $\mathscr{P}$ has a limit point x. Therefore, $x_k \to x$. □

4.5 Archimedean Property

Theorem 5 (**Arch**: Archimedean Property) $k \to +\infty$*, i.e., for every* $u \in \mathbb{R}$ *we have* $k \geq u$ *for large enough* k*.*

As usual this implies that the set $\mathbb{Q}$ of rational numbers is dense, i.e., each interval $[u, v]$ contains a rational number.

Proof Consider the sequence (k) of natural numbers in $[a, b] = [0, +\infty]$ and let $[u, v] \in \mathscr{P}$ be defined by: "$[u, v]$ contains all large enough integers". This property is obviously global. The assertion "$[u, v] \in \mathscr{P}$" means that v is greater than all the integers, and that u is not greater than all the integers; thus $\mathscr{P}$ is subtractive (see Example 3). By the GL axiom, $\mathscr{P}$ has a limit point x, i.e., $k \to x$. If x were finite, we could find an interval of the type $[v - 1, v]$ (where v is finite) which contains all integers $\geq k$. For such a k, we have $v - 1 < k \implies v < k + 1$, which is impossible. Therefore, $x = +\infty$. □

An alternate proof uses the GL axiom and the property defined by "$[u, v]$ contains infinitely many integers".

4.6 Cauchy Sequences

Instead of the classical definition of a Cauchy sequence using double indexing, we feel that the following definition is somehow simpler.

Definition 16 A sequence (x_k) is *Cauchy* if for all $\varepsilon > 0$, we have, starting from a certain index K, the inequality $|x_K - x_k| < \varepsilon$ for all $k \geq K$.

In other words, the sequence is eventually "almost stationary". This definition, of course, requires that the x_k are eventually all finite. The usual definition of the usual convergence (towards a finite limit x) replaces x_K by x in the above inequality. Since $|x_K - x_k| \leq |x_K - x| + |x - x_k|$, any convergent sequence is a Cauchy sequence.

Remark 1 The classical definition is: for all $\varepsilon > 0$, we have $|x_\ell - x_k| < \varepsilon$ for all large enough k and ℓ ($\geq K$). Since $|x_\ell - x_k| \leq |x_\ell - x_K| + |x_K - x_k|$ this is equivalent to the above definition.

Remark 2 A Cauchy sequence is bounded, because for any given $\varepsilon > 0$, we have $|x_k| < \varepsilon + |x_K|$ for all $k \geq K$.

The following *Cauchy criterion* is Bolzano's theorem:

Theorem 6 (**Cauchy**: Cauchy's Completeness) *A sequence is convergent if and only if it is a Cauchy sequence.*

Proof It is enough to show that a Cauchy sequence (x_k) converges. Property $\mathscr{P}$ defined by "$[u, v]$ contains every x_k for all large enough k". is obviously global. It is also o-subtractive, because if $t < u < v < w$ and $[t, w]$ satisfies $\mathscr{P}$, it is impossible for both $[t, u]$ and $[v, w]$ to contain x_k for infinitely many values of k, since that would contradict the Cauchy property $|x_\ell - x_k| < \varepsilon$ for $\varepsilon = (v - u)$. Hence either $[t, v] \in \mathscr{P}$ or $[u, w] \in \mathscr{P}$. By the GL axiom, $\mathscr{P}$ has a limit point x, that is, $x_k \to x$. The limit x is finite because the sequence (x_k) is bounded. □

4.7 Nested Intervals of Cantor and Adjacent Sequences

We consider families of intervals $[r, s]$ belonging to $[a, b]$. Such families do not have a common point if their intersection is empty.

Theorem 7 (**Cantor**) *Each family of intervals* $[r, s]$ *with no common point admits a finite subfamily with no common point.*

Proof Let $\mathscr{S}$ be the family of intervals $[r, s]$ with no common point:

$$\bigcap_{[r,s]\in\mathscr{S}} [r, s] = \emptyset.$$

Let $\mathscr{P}$ be the property that there is a finite sub-family of $\mathscr{S}$ with no common point in $[u, v]$. This property is local. Indeed, no point x belongs to all intervals of $\mathscr{S}$, so there exists an interval $[r, s] \in \mathscr{S}$ which does not contain x. It is possible to chose a neighborhood $V(x)$ disjoint from this interval $[r, s]$. Each $[u, v]$ adapted to $V(x)$ then satisfies $\mathscr{P}$, since it does not contain any point of $[r, s]$ (which by itself constitutes a finite sub-family of $\mathscr{S}$). The property $\mathscr{P}$ is also additive, because given $u < v < w$, if $[u, v]$ does not contain a common point of a finite sub-family of $\mathscr{S}$, and if $[v, w]$

does not contain a common point of another finite sub-family of $\mathscr{S}$, then $[u, v]$ does not contain a common point of the union of the two finite sub-families. By the LG axiom, $[a, b] \in \mathscr{P}$, that is, there exists a finite sub-family of $\mathscr{S}$ without any common point in $[a, b]$. □

Theorem 8 (**Nest**: Cantor's Property) *Every sequence* $[r_k, s_k]$ *of nested intervals (such that* $[r_{k+1}, s_{k+1}] \subset [r_k, s_k]$ *for all k) has a common point (common to each of the intervals).*

This is an immediate consequence of Cantor's theorem, since all finite sub-families of nested intervals sequences $[r_k, s_k]$ has as an intersection in the last (smallest) interval, which is not-empty. As a consequence the sequence has a common point. However it is instructive to show a direct proof using the GL axiom:

Proof Let $\mathscr{P}$ be the property that $[u, v]$ contains one of the intervals $[r_k, s_k]$ (hence all intervals for large enough k). The property $\mathscr{P}$ is clearly global. If it is *not* subtractive, there exists $u < v < w$ such that $[u, w] \in \mathscr{P}$ with $r_k < v < s_k$ for all k large enough: v is then a common point for all such intervals. Otherwise, $\mathscr{P}$ is subtractive, and by the GL axiom, $\mathscr{P}$ has a limit point x. This point is necessarily common to the intervals $[r_k, s_k]$, otherwise we could find an interval $[u, v]$ which contains x and disjoint from an interval $[r_k, s_k]$, which contradicts that $\mathscr{P}$ has a limit at x. □

An easy consequence is

Theorem 9 (Cantor) $[a, b]$ *is uncountable.*

Proof We may assume that $[a, b]$ is bounded. If it is countable, let us write $[a, b] = \{x_1, x_2, \ldots, x_k, \ldots\}$. Set $[a_0, b_0] = [a, b]$. For all integer $k \geq 0$, define by induction a subinterval $[a_{k+1}, b_{k+1}]$ of $[a_k, b_k]$ which does not contain x_k (by example, partition $[a_k, b_k]$ into three sub-intervals of the same length, and define $[a_{k+1}, b_{k+1}]$ as the first sub-interval of three that does not contain x_k). The sequence of intervals $[a_k, b_k]$ has a common point $x \in [a, b]$ distinct from all of the $x_1, x_2, \ldots, x_k, \ldots$, which is impossible. □

Definition 17 Two sequences (r_k), (s_k) in $[a, b]$ are *adjacent* if (r_k) is nondecreasing, (s_k) is nonincreasing, and $s_k - r_k$ tends to 0.

Theorem 10 (Adjacent Sequences) *Two adjacent sequences converge to the same limit.*

Proof The difference $s_k - r_k$ decreases since r_k increases and s_k decreases. As it tends to zero, it is always ≥ 0. So $r_k \leq s_k$ for every k, and the intervals $[r_k, s_k]$ are nested. Let x be a common point in these intervals: $r_k \leq x \leq s_k$ for every k. Since the width of $[r_k, s_k]$ tends to zero, each neighborhood $V(x)$ contains $[r_k, s_k]$ for large enough k. Hence both sequences tend to x. □

For completeness we observe the following.

Theorem 11 *The LG axiom is equivalent to the two theorems of the adjacent sequences Theorem 10 and of Archimedes Theorem 5.*

Proof It is enough to show that both theorems imply the GL axiom. Let $\mathscr{P}$ be global and subtractive. One proceeds by dichotomy. Let $[a_0, b_0] = [a, b]$ unless $[a, b] = [-\infty, +\infty]$, in which case we can define $[a_0, b_0]$ to be equal to the first of the two intervals $[-\infty, 0]$ or $[0, +\infty]$ which satisfies $\mathscr{P}$. Thus we can always assume that at least one of the two interval extremities are finite. We define them by induction $[a_{k+1}, b_{k+1}]$ equal the first of the two following intervals that satisfy $\mathscr{P}$:

- $[a_k, (a_k + b_k)/2]$ or $[(a_k + b_k)/2, b_k]$, if a_k and b_k are finite;
- $[a_k, a_k + 1]$ or $[a_k + 1, +\infty]$, if $b_k = +\infty$ (a_k being finite);
- $[-\infty, b_k - 1]$ or $[b_k - 1, b_k]$, $a_k = -\infty$ (b_k being finite).

As $\mathscr{P}$ is subtractive, this sequence is well defined. Since (a_k) is nondecreasing, and (b_k) is nonincreasing, we have three cases to consider:

- a_k and b_k are finite for large enough k; then $b_k - a_k = 2^{-k}(b - a)$ tends to 0 since (by Theorem 5) $2^k > k \to +\infty$; the sequences (a_k) and (b_k) are adjacent, so by Theorem 10 they converge to the same limit.
- $b_k = +\infty$ for every k; we then have $a_0 = 0$ and $a_{k+1} = a_k + 1$ for any $k \geq 0$, where $a_k = k \to \infty$ by Theorem 5.
- $a_k = -\infty$ for every k; we then have the same $b_k = -k \to -\infty$ by Theorem 5.

In all cases, a_k and b_k tend to the same limit x (finite or infinite). Each neighborhood $V(x)$ for every k large enough, contains the interval $[a_k, b_k] \in \mathscr{P}$. Therefore $\mathscr{P}$ has a limit point x. □

4.8 Bolzano–Weierstrass Property

Definition 18 A *limit point* x of a set E (also called *accumulation point*) is such that each neighborhood $V(x)$ contains infinitely many points of E. That is to say, the property that $[u, v]$ contains infinitely many points of E has x as a limit point in the sense of Definition 11.

A *limit point* x of a sequence (x_k) (also called *cluster point*, or *adherent value*) is such that each neighborhood $V(x)$ contains x_k for infinitely many values of k. That is to say, the property that $[u, v]$ contains x_k for infinitely many values of k has x as a limit point in the sense of Definition 11.

Theorem 12 (**BW**: Bolzano–Weierstrass) *Any* infinite *set of points has a limit point.*

If the set is bounded, this limit point is finite.

Proof Let E be an infinite set of points, and $\mathscr{P}$ be the property with the interval $[u, v]$ containing infinitely many points in E. This property is global by hypothesis. It is also subtractive: if $[u, w]$ contains infinitely many points in E, at least one of those sub-intervals $[u, v]$, $[v, w]$ must have infinitely many points in E. By the GL axiom, $\mathscr{P}$ has a limit point x, i.e., x is a limit point of E. □

Note that by contraposition, any locally finite set is finite. Also, by the same argument, any uncountable set of points has a *condensation point* (that is, such that every neighborhood of it contains uncountably many points of E).

Theorem 13 (**BW** for Sequences) *Any sequence has a limit point.*

This is a consequence of the preceding theorem applied to the set of the sequence values, if one considers the two cases where the set is finite or infinite. A direct proof using the GL axiom is as follows.

Proof Let (x_k) be any point sequence and let $\mathscr{P}$ be the property that an interval $[u, v]$ contains x_k for infinitely many values of k. The property is evidently global. It is also subtractive: if $[u, w]$ contains x_k for infinitely many values of k, at least one of $[u, v]$, $[v, w]$ has the same property. By the GL axiom, $\mathscr{P}$ has a limit point x, that is, x is a limit point (in the usual sense of Definition 18) of the sequence (x_k). □

4.9 Heine–Borel–Lebesgue Cover

Recall that a family of intervals *cover* a set of points E if each point in E is in at least one of the intervals of that family. For example, $\mathscr{R} = \{]r_i, s_i[\}_{i\in I}$ covers E if $E \subset \bigcup_{i\in I}]r_i, s_i[$. We also say that this family is a *cover* of E. In addition, if this family is composed of a finite number of intervals, then it is a *finite cover* of E.

Theorem 14 *(***BL**: Borel–Lebesgue (sometimes known as Heine–Borel)) *Any cover of* $[a, b]$ *by means of open intervals admits a finite subcover.*

In other words, given any family $\mathscr{R}$ of open intervals covering $[a, b]$, we can find a finite number of intervals $]r_k, s_k[$ $(k = 1, \ldots, m)$ of $\mathscr{R}$ such that each point in $[a, b]$ belongs to at least one of $]r_k, s_k[$:

$$[a, b] \subset \bigcup_{k=1}^{m}]r_k, s_k[.$$

Proof Let $\mathscr{P}$ be the property that the interval $[u, v]$ is covered by a *finite* number of intervals $]r, s[$ of $\mathscr{R}$. As $\mathscr{R}$ covers $[a, b]$, each $x \in [a, b]$ belongs to some interval $]r, s[$ of $\mathscr{R}$. Taking $V(x) =]r, s[$ as a neighborhood of x, every interval $[u, v] \subset V(x)$ is covered by a finite number (equal to 1) of intervals of $\mathscr{R}$, that is, $]r, s[$ itself. Hence $\mathscr{P}$ is local. It is also additive: if $[u, v]$ and $[v, w]$ covered each one by a finite number of open intervals of $\mathscr{R}$, their union is a finite cover of $[u, w]$. By the LG axiom, $\mathscr{P}$ is global, i.e., satisfied by $[a, b]$. □

4.10 Cousin Partition

Definition 19 A *partition* (of intervals) of $[a, b]$ is a finite cover of $[a, b]$ by non-overlapping intervals $[u, v] \subset [a, b]$.

In other words, a partition of $[a, b]$ corresponds to a subdivision, that is, a finite number of points $a = u_1 < u_2 < \cdots < u_m = b$, such that the intervals $[u_k, u_{k+1}]$ $(1 \leq i < m)$ do not overlap and cover $[a, b]$.

Definition 20 An *environment* V of a set of points E is a family of neighborhoods, with one neighborhood $V(x)$ for every point $x \in E$.

Definition 21 An interval $[u, v]$ is *adapted* to the environment V if there exists $x \in [u, v]$ such that $[u, v] \subset V(x)$. A set of intervals is *adapted* to the environment V if every interval is.

Thus, a partition $\pi = \{[u_i, u_{i+1}]\}_{1 \leq i < m}$ of $[a, b]$ is *adapted* to the environment V (on $[a, b]$) if it exists for each of the intervals of which it is composed, that is, if for all i $(1 \leq i < m)$, there exists $x \in [u_i, u_{i+1}] \subset V(x)$.

Theorem 15 (**Cousin**: Cousin's Partition) *For any environment V of $[a, b]$, there exists a partition of $[a, b]$ adapted to V.*

Proof Let $\mathscr{P}$ be the property that the interval $[u, v]$ admits a partition adapted to V. For all $x \in [a, b]$, every adapted interval $[u, v]$ to $V(x)$ is for itself an adapted partition of $[u, v]$ to V. Hence $\mathscr{P}$ is local. It is also additive: if $[u, v]$ and $[v, w]$ admit each one a partition adapted to V, the union of the two partitions constitute a partition of $[u, w]$ adapted to V. By the LG axiom, $\mathscr{P}$ is global. □

For completeness we observe the following.

Proposition 8 *The LG axiom is equivalent to Cousin's theorem.*

Proof It is enough to show that Cousin's theorem implies the LG axiom. Let $\mathscr{P}$ be a local and additive property in $[a, b]$. As $\mathscr{P}$ is local, there exists an environment V such that every adapted interval to V satisfies $\mathscr{P}$. A Cousin's partition corresponding to V is then composed of intervals that satisfy $\mathscr{P}$. Since $\mathscr{P}$ is additive, it follows that $[a, b] \in \mathscr{P}$: $\mathscr{P}$ is global. □

5 Equivalence Between Completeness Axioms

In this section we prove the equivalence between the various completeness axioms. This of course is not required in an elementary course but is satisfactory for logical consistency.

Proposition 9 *Local-global (LG or GL) axioms are equivalent to any of the following statements:*

1. *Existence of a Dedekind cut point;*
2. *Existence of supremum (or infimum);*
3. *Principle of continuous induction;*
4. *Intersection theorems of Cantor + Archimedean property;*
5. *Nested intervals theorem + Archimedean property;*
6. *Adjacent sequences theorem + Archimedean property;*
7. *Monotone limit theorem;*
8. *Cauchy's criterion + Archimedean property;*
9. *Bolzano–Weierstrass theorem (for sets or for sequences);*
10. *Heine–Borel–Lebesgue covering theorem;*
11. *Existence of Cousin's partition.*

Proof We have already directly proven each one of the results from LG or GL axioms. The converse proofs given above show the equivalences with the statements 1, 2, 3, 6, and 11 . Furthermore, the implications $4 \implies 5 \implies 6$ have already been seen, hence the equivalences with statements 4 and 5.

We can conclude with the following implications: ($9 \implies 7 \implies 6$), ($8 \implies 6$), and ($10 \implies 11$).

$9 \implies 7$: We have seen that the Bolzano–Weierstrass theorem for sets implies that for sequences. We now show that the Bolzano–Weierstrass theorem for sequences implies the monotone convergence theorem. Let (x_k) be a monotonic sequence and $V(x)$ an interval which is a neighborhood of x. The sequence (x_k) admits a limit point x such that $x_\ell \in V(x)$ for infinitely many values of ℓ. Let K be such that $x_K \in V(x)$ and $k \geq K$. There exists an index $\ell > k$ such that $x_\ell \in V(x)$. As this sequence is monotonic, x_k lies between x_K and x_ℓ, hence $x_k \in V(x)$ for all $k \geq K$, which shows that x is the limit of (x_k).

$7 \implies 6$: If two adjacent sequences (r_k), (s_k) are monotone, then they tend to limits: $r_k \to r$ and $s_k \to s$. Since $s_k - r_k \to 0$, we deduce $s - r = 0$, hence $r = s$, which proves the adjacent sequences theorem.
The sequence $x_k = k$ is increasing and tends to a limit x. If x is finite, we have both $k \to x$ and $k + 1 \to x + 1$ so $x = x + 1$ which is impossible. Hence $x = +\infty$, which proves the Archimedes property.

$8 \implies 6$: It is enough to show that the Cauchy criterion implies the adjacent sequences theorem. Two adjacent sequences (r_k), (s_k) are such that $s_k - r_k$ is nonincreasing and tends to 0. For all $\varepsilon > 0$, we have then $s_K - s_k \leq s_K - r_K < \varepsilon$ and $r_k - r_K \leq s_K - r_K < \varepsilon$ for all $k \geq K$. These are Cauchy sequences, hence converge: $r_k \to r$ and $s_k \to s$. As $s_k - r_k \to 0$, we deduce that $s - r = 0$ or $r = s$.

10 $\implies$ 11: Let V be an environment of $[a, b]$ which we assume is of open intervals. This constitutes a covering of $[a, b]$. Extract a finite covering $V(x_k) =]r_k, s_k[$ for $1 \le k \le K$. Relabelling if necessary, we can always assume that $x_1 < x_2 < \cdots < x_K$ and assume that K is minimal. Pick x_{i_1} such that $a \in V(x_{i_1})$ and as long as $b \notin V(x_{i_j})$ define a finite sequence x_{i_j} such that $x_{i_{j+1}} > x_{i_j}$ and $V(x_{i_{j+1}})$ overlaps with $V(x_{i_j})$, until $V(x_{i_{m-1}}) \ni b$. One obtains a finite covering $V(x_{i_k}) =]r_{i_k}, s_{i_k}[$ for $1 \le k < m$ where each $V(x_{i_k})$ overlaps with $V(x_{i_{k+1}})$. We can then choose m points $a = u_1 < u_2 < \cdots < u_m = b$ with $u_{k+1} \in V(x_{i_k}) \cap V(x_{i_{k+1}})$. Every $[u_k, u_{k+1}]$ is then adapted to $V(x_{i_k})$ for $1 \le k < m$. These intervals forms an adapted Cousin partition to V.

□

6 Elementary Theorems of Real Analysis

6.1 *Continuous Functions*

We consider functions defined on $[a, b]$, with values in $\mathbb{R}$ or $\mathbb{C}$, or more generally in $\mathbb{R}^n$ or $\mathbb{C}^n$, or every vector space of finite dimension over $\mathbb{R}$ or $\mathbb{C}$, or even more generally, in a Banach space over $\mathbb{R}$ or $\mathbb{C}$. We note $|\cdot|$ the corresponding absolute value, modulus, or norm.

We advocate the following definition of continuity.

Definition 22 A function f is *continuous* in a point x if for all $\varepsilon > 0$, the property "$|f(v) - f(u)| < \varepsilon$" is local in x. A function f is *continuous* on a set E if it is continuous in each point of E.

In other words, f is continuous at x if for all $\varepsilon > 0$, there exists a neighborhood $V(x)$ in which $u \le x \le v$ implies $|f(u) - f(v)| < \varepsilon$. When $x = b$, it is a continuity to the left at b; when $x = a$, it is a continuity to the right at a.

Conversely, f is discontinuous at x if there exists $\varepsilon > 0$ for which the property "$|f(v) - f(u)| \ge \varepsilon$" is not limit at x.

Remark 3 The continuity definition in a point x is equivalent to the classical definition: for all $\varepsilon > 0$, there exists a neighborhood $V(x)$ such that for all $t \in V(x)$, $|f(t) - f(x)| < \varepsilon$. Indeed we obtain this condition from the definition above when getting $[u, v] = [x, t]$ if $t \ge x$, $[u, v] = [t, x]$ otherwise. Conversely, if $[u, v]$ is adapted to $V(x)$, we have $|f(v) - f(u)| \le |f(v) - f(x)| + |f(u) - f(x)| < 2\varepsilon$.

Remark 4 Saying that f is continuous at x is the same as saying that $f(t)$ tends toward $f(x)$ when $t \to x$. When $x = b$, this is a limit to the left at b, denoted $f(b^-)$; when $x = a$, this is a limit to the right at a, denoted $f(a^+)$.

It is possible for f to only be defined on $]a, b[$ but having finite limits $f(a^+)$ and $f(b^-)$. We can say then that f is *continuous* on $[a, b]$, in the sense where we can extend by continuity f by setting $f(a) = f(a^+)$ and $f(b) = f(b^-)$.

Example 12 The function $f(x) = \dfrac{1}{1+x^2}$ is continuous on $[-\infty, +\infty]$ (extend by continuity setting $f(\pm\infty) = 0$).

Example 13 The function $f(x) = \arctan(x)$ is continuous on $[-\infty, +\infty]$ (extend by continuity setting $f(-\infty) = -\pi/2$ and $f(+\infty) = \pi/2$).

Example 14 The function $f(x) = x^2$ is continuous on any bounded interval. But with our definition[7] it is not continuous (cannot be extended by continuity) on $[-\infty, +\infty]$.

Theorem 16 (**BT**: Boundedness Theorem) *Each continuous function on* $[a, b]$ *is bounded on* $[a, b]$.

Notice that $[a, b]$ may very well be unbounded (a and/or b can be infinite).

Proof The property $\mathscr{P}$ defined by "f is bounded on $[u, v]$" is local since for given $\varepsilon > 0$, any u in the neighborhood $V(x)$ satisfies $|f(u)| \leq |f(u) - f(x)| + |f(x)| \leq |f(x)| + \varepsilon$. The property $\mathscr{P}$ is also additive, because if $|f| \leq M$ on $[u, v]$ and $|f| \leq M'$ on $[v, w]$ then $|f| \leq \max(M, M')$ on $[u, w]$. By the LG axiom, f is bounded on $[a, b]$. □

The above proof shows that, more generally, any *locally bounded* function is (globally) bounded.

Theorem 17 (**EVT**: Extreme Value Theorem) *Each continuous real function f on $[a, b]$ reaches its maximum, i.e., there exists $x \in [a, b]$ such that $f(x) \geq f(t)$ for all $t \in [a, b]$.*

Of course, considering $-f$, each real continuous function f on $[a, b]$ reaches its minimum.

Proof The property $\mathscr{P}$: "there does not exist a value of f which is greater than any value of f on $[u, v]$" (in other words $\forall t \in [a, b],\ \exists x \in [u, v],\ f(t) \leq f(x)$) is clearly global. It is also subtractive, otherwise the greater of the two values of f is greater than f on $[u, w]$. By the GL axiom, $\mathscr{P}$ has a limit point x. We have then, by continuity, $\forall t \in [a, b]$, $f(t) < f(x) + \varepsilon$ for all $\varepsilon > 0$, hence $f(t) \leq f(x)$ for all $t \in [a, b]$. □

Interestingly, the proof extends verbatim to the more general case where f is *upper semi-continuous*. Also notice that the boundedness theorem was not required for this proof, which, therefore, provides another proof for boundedness since $\min f \leq f \leq \max f$.

[7] We found it convenient for later developments that functions assume only finite values in order to leverage on the complete metric space property of the set of function values.

Theorem 18 (**IVT**: Intermediate Value Theorem) *Each continuous real function on* $[a, b]$ *takes any value between* $f(a)$ *and* $f(b)$.

In other words, if y is any value taken between $f(a)$ and $f(b)$, there exists x such that $f(x) = y$.

Proof The property $[u, v] \in \mathscr{P}$ defined by "y is between $f(u)$ and $f(v)$" (that is, "$f(u) \leq y \leq f(v)$ or $f(v) \leq y \leq f(u)$") is global. It is also subtractive, because if $u < v < w$ and y is between $f(u)$ and $f(w)$, then whatever the value of $f(v)$, y is either between $f(u)$ and $f(v)$, or between $f(v)$ and $f(w)$. By the GL axiom, $\mathscr{P}$ has a limit point x. By continuity, for all $\varepsilon > 0$, we have $f(x) - \varepsilon < y < f(x) + \varepsilon$, hence $y = f(x)$. □

Theorem 19 (**Heine**'s Theorem) *Every continuous function of a bounded interval* $[a, b]$ *is uniformly continuous.*

Proof Let f be continuous on $[a, b]$ and $\varepsilon > 0$. The property $\mathscr{P}$: "$|f(x) - f(y)| < \varepsilon$ for all x, y sufficiently close in $[u, v]$" (that is, there exists $\delta > 0$ such that $|x - y| < \delta \implies |f(x) - f(y)| < \varepsilon$) is local by definition of continuity. It is also o-additive: if $t < u < v < w$ and if for all x, y sufficiently close in $[t, v]$ or in $[u, w]$ one has $|f(x) - f(y)| < \varepsilon$, then all $x, y \in [t, w]$ such that $|x - y| < v - u$ will be both in $[t, v]$ or both in $[u, w]$, so that we will always have $|f(x) - f(y)| < \varepsilon$ for sufficiently close x, y. By the LG axiom, $\mathscr{P}$ is global, which means that f is uniformly continuous on $[a, b]$. □

6.2 *Differentiable Functions*

We advocate the following definition of differentiability.

Definition 23 A function f is *differentiable* at a finite point x with derivative $f'(x) = \lambda$ if for all $\varepsilon > 0$, the property $[u, v] \in \mathscr{P}$ defined by

$$|f(v) - f(u) - \lambda(v - u)| < \varepsilon(v - u)$$

is local in x. A function f is *differentiable* on a (bounded) set E if it is differentiable at all points of E which defines the *derivative* f' of f on E.

In other words, f has a derivative $\lambda = f'(x)$ at x if for all $\varepsilon > 0$, there exists a neighborhood $V(x)$ in which $u \leq x \leq v$ implies $|f(v) - f(u) - f'(x) \cdot (v - u)| < \varepsilon(v - u)$. If f is defined on $[a, b]$ and a or b are finite, we can have $x = a$ or $x = b$ in the definition of the derivative; this is then a derivative to the left at b if $x = b$, a derivative to the right at a if $x = a$.

Notice that if f takes values in a vector space, $\lambda \cdot (v - u)$ is the vector product λ by the scalar $(v - u)$, and the derivative $f'(x)$ has vector values.

Remark 5 For real-valued functions the above definition of derivative at x point is equivalent to the classical definition:

$$\frac{f(x) - f(t)}{x - t} \to \lambda$$

where t tends to x. Indeed, we get this condition from the definition above by taking $[u, v] = [x, t]$ if $t \geq x$, $[u, v] = [t, x]$ otherwise, and dividing by $(v - u)$. Conversely, if $[u, v]$ is adapted to $V(x)$, we have $|f(v) - f(u) - \lambda(v - u)| = |f(v) - f(x) - \lambda(v - x) + f(x) - f(u) - \lambda(x - u)| \leq \varepsilon\big(|v - x| + |x - u|\big) = \varepsilon(v - u)$. (This is sometimes referred to in the literature as the "straddle lemma").

Proposition 10 *Every differentiable function is continuous.*

Proof If f is differentiable with derivative λ at x, the property $|f(v) - f(u)| < (|\lambda| + \varepsilon)(v - u)$ is local at x. But if $[u, v]$ is small enough, for a given $\varepsilon' > 0$, $(|\lambda| + \varepsilon)(v - u) < \varepsilon'$, hence the property $|f(v) - f(u)| < \varepsilon'$ is local at x. □

Inspired by Cohen and Bers [3, 6], we advocate the use of the following theorem in place of the classical mean value theorem (see Remark 6 below).

Theorem 20 (Finite Increase Inequality[8]) *Let f, g be two differentiable functions on $[a, b]$, where g assumes real values.*

- *if $|f'| < g'$ on $[a, b]$ then $|f(b) - f(a)| < g(b) - g(a)$;*
- *if $|f'| \leq g'$ on $[a, b]$ then $|f(b) - f(a)| \leq g(b) - g(a)$;*

Proof Let $x \in [a, b]$, suppose $|f'(x)| < g(x)$ and let $\varepsilon > 0$ small enough such that $|f'(x)| + \varepsilon < g'(x) - \varepsilon$.

For $u \leq x \leq v$ in a neighborhood $V(x)$, one has $|f(v) - f(u)| < |f'x)| \cdot (v - u) + \varepsilon(v - u) < g'(x) \cdot (v - u) - \varepsilon(v - u) < g(v) - g(u)$. This implies that the property "$|f(v) - f(u)| < g(v) - g(u)$" is local. This property is also additive for if $u < v < w$, $|f(v) - f(u)| < g(v) - g(u)$ and $|f(w) - f(v)| < g(w) - g(v)$ imply $|f(w) - f(u)| \leq |f(w) - f(v)| + |f(v) - f(u)| < g(w) - g(v) + g(v) - g(u) = g(w) - g(u)$. By the LG axiom, the property is global: $\big|f(b) - f(a)\big| < g(b) - g(a)$.

For the second part, we can replace $g(x)$ by $g(x) + \varepsilon x$ for $\varepsilon > 0$, so that $|f'| < g' + \varepsilon$, hence $|f(b) - f(a)| < g(b) - g(a) + \varepsilon(b - a)$. Since $\varepsilon > 0$ is arbitrarily small, we have $|f(b) - f(a)| \leq g(b) - g(a)$. □

An alternate proof for the second part considers the property "$|f(v) - f(u)| \leq g(v) - g(u) + \varepsilon(v - u)$".

The above theorem is enough to prove the following important results:

- if $f' = 0$ on $[a, b]$ then f is constant there (take g constant); thus an antiderivative is unique up to an additive constant;

[8] A literal translation of the French "inégalité des accroissements finis", which advantageous replaces the "théorème des accroissements finis", which is the mean value theorem.

- if $g' > 0$ on $[a, b]$ then g is increasing; if $g' \geq 0$ on $[a, b]$ then g is nondecreasing (take f constant).

Remark 6 Cohen and Bers wrote:

> *With characteristic vigor, L. Bers announced in a recent conversation: "Who needs the mean value theorem! All we want as a start in elementary calculus is the proposition that if $f'(x) = 0$ for all x in $[a, b]$, then f is constant."* [6]
>
> *The "full" mean value theorem [...] is a curiosity. It may be discussed together with another curiosity, Darboux' theorem that every derivative obeys the intermediate value theorem.* [3]

The actual "mean value theorem", which states that there exists $x \in \;]a, b[$ such that

$$\frac{f(b) - f(a)}{b - a} = f'(x).$$

gives no indication of the position of point x in the interval $[a, b]$. This proof shows that the only condition is that $\frac{f(b)-f(a)}{b-a}$ lies between two possible values of the derivative, and as such, is related to Darboux's theorem that f' takes any value between $f'(a)$ and $f'(b)$ on $[a, b]$. We couldn't find a direct and easy proof of the mean value theorem using the LG or GL axiom. Darboux's theorem can be proved using the fact that the derivative f' vanishes at an extremum of f, and the mean value theorem then becomes an easy consequence of Darboux's theorem: if f' does not take the value $\lambda = \frac{f(b)-f(a)}{b-a}$ then by Darboux's theorem f' is either always greater or always less, which by the finite increase inequality implies either $\frac{f(b)-f(a)}{b-a} > \lambda$ or $< \lambda$, a contradiction.

7 Conclusion and Perspectives

Our objective is twofold. First we would like to draw attention to the local-global principle as a new efficient and enjoyable tool for proving the basic theorems of real analysis. Second, we aim to clarify the local-global concept to possibly improve the teaching of real analysis at undergraduate and graduate levels.

As a future work the LG/GL concept may be used as a basis for a new presentation of the integral, just as Cousin's lemma was used to build the Kurzweil–Henstock integral [12, 18, 23]. In such an approach the so-called "fundamental theorem of calculus", appropriately generalized, can become the actual definition for a novel notion of the antiderivative function.

Acknowledgements The authors would like to thank Pete Clark for his detailed review of the paper and his many useful comments and suggestions.

References

1. Andre, N.R., M.E. Susannah and E.P. Adam. 2013. An Analysis of the First Proofs of the Heine-Borel Theorem. Convergence, vol. 10 Mathematical Association of America.
2. Bartle, R.G. 1964, 1976. *The Elements of Real Analysis*. Wiley.
3. Bers, L. 1967. On avoiding the mean value theorem. *American Mathematical Monthly* 74: 5, 583.
4. Bottazzini, U. 1986. *The Higher Calculus: A History of Real and Complex Analysis from Euler to Weierstrass*. Springer.
5. Clark, P.L. 2011. The instructor's guide to real induction. Preprint.
6. Cohen, L.W. 1967. On being mean to the mean value theorem. *American Mathematical Monthly* 74 (5): 581–582.
7. Courant, R., and F. John. 1965. *Introduction to Calculus and Analysis I*. Wiley. Springer (1989).
8. Dedekind, R. 1901. *Essays on the theory of numbers-Continuity and Irrational Numbers – The Nature and Meaning of Numbers*. The Open Court Publishing Company. Dover Publications (1963, 2010).
9. Dieudonné, J. 1960, 1969. *Foundations of Modern Analysis*. Academic Press.
10. Duren, W.L. 1957. Mathematical induction in sets. *American Mathematical Monthly* 64 (8): 19–22.
11. Ford, L.R. 1957. Interval-additive propositions. *American Mathematical Monthly* 64 (2): 106–108.
12. Gordon, R.A. 1994. The Integrals of Lebesgue, Denjoy, Perron, and Henstock. Graduate Studies in Mathematics, vol. 4, American Mathematical Society.
13. Gordon, R.A. 1998. The use of tagged partitions in elementary real analysis. *American Mathematical Monthly* 105 (2): 107–117.
14. Grabiner, J.V. 1983. Who gave you the epsilon? Cauchy and the origins of rigorous calculus. *American Mathematical Monthly* 91: 185–194.
15. Guedes de Figueiredo, D. 1973, 1996. Análise I, LTC Editora.
16. Guyou, C. 1946. *Algèbre et Analyse à l'Usage des Candidats aux Grandes Écoles*. Paris: Vuibert.
17. Hathaway, D. 2011. Using continuity induction. *The College Mathematics Journal* 42 (3): 240–242.
18. Henstock, R. 1963. *Theory of Integration*. Butterworths.
19. Jungck, G. 1966. Interval induction. *American Mathematical Monthly* 73 (3): 295–297.
20. Kalantari, I. 2007. Induction over the continuum. In *Induction, Algorithmic Learning Theory, and Philosophy*, chap. 5, pp. 145–154. Springer.
21. Kimber, J.E. 1965. Two extended Bolzano-Weierstrass theorems. *American Mathematical Monthly* 72 (9): 1007–1012.
22. Klein, F. 1895. Ueber Arithmetisirung der Mathematik. In *Göttinger Nachrichten* (Geschäftliche Mittheilungen), p. 82.
23. Kurzweil, J. 1957. Generalized ordinary differential equations and continuous dependence on a parameter. *Czechoslovak Mathematical Journal* 7 (3): 418–449.
24. Lages Lima, E. 1989. Análise Real I. Coleção Matemática Universitária.
25. Lang, S. 1968. *Analysis I*. Addison-Wesley, Springer (1983, 1997).
26. Leinfelder, H. 1982. A unifying principle in real analysis. *Real Analysis Exchange* 8: 511–518.
27. Lützen, J. 2003. The foundation of analysis in the 19th century. In *A History of Analysis*, chap. 6. Edited by Jahnke, H.N., pp. 155–195. American Mathematical Society.
28. Moss, R., and G. Roberts. 1968. A creeping lemma. *American Mathematical Monthly* 75: 649–652.
29. Rudin, W. 1953, 1964, 1976. *Principles of Mathematical Analysis*. McGraw-Hill.
30. Shanahan, P. 1972. A unified proof of several basic theorems of real analysis. *American Mathematical Monthly* 79 (8): 895–898.

Quantitative Logic Reasoning

Marcelo Finger

Abstract In this paper we show several similarities among logic systems that deal simultaneously with deductive and quantitative inference. We claim it is appropriate to call the tasks those systems perform as Quantitative Logic Reasoning. Analogous properties hold throughout that class, for whose members there exists a set of linear algebraic techniques applicable in the study of satisfiability decision problems. In this presentation, we consider as Quantitative Logic Reasoning the tasks performed by propositional Probabilistic Logic; first-order logic with counting quantifiers over a fragment containing unary and limited binary predicates; and propositional Łukasiewicz Infinitely-valued Probabilistic Logic.

1 Introduction

Quantitative Logic Reasoning aims at providing a unified treatment to several tasks that involve both a deductive logic reasoning and some form of inference about quantities. Typically, reasoning with quantities involves probabilities and/or cardinality assessments. Superficially, we are dealing with such distinct quantitative inferential capabilities but it is our aim to clarify that, to some significant extent, these approaches share a considerable set of common features, which include, but are not restricted to:

- similar reasoning tasks with quantities, which typically involve decision problems such as satisfiability or entailment assessments;
- similarly structured fragments that lead to the existence of normal forms;
- similar characterizations of consistency in terms of coherence;
- similar formulations based on Linear Algebra;

M. Finger (✉)
Department of Computer Science, University of São Paulo, Sao Paulo, Brazil
e-mail: mfinger@ime.usp.br

W. Carnielli and J. Malinowski (eds.), *Contradictions, from Consistency to Inconsistency*, Trends in Logic 47, https://doi.org/10.1007/978-3-319-98797-2_12

- similar decision algorithms employing SAT-based column generation;
- similar complexity of decision problems, which for the fragments covered in this work are "only" NP-complete.

We believe that the presence of such similarities elicits the grouping of several logic systems under the name of Quantitative Logic Reasoning systems.

We explore the shared properties of three logic systems with the aim of bringing forward the similarities as well as the particularities of each system. For that, we present some well known results, which are employed as a basis for the development of quantitative reasoning techniques; we also present original results, mainly in dealing with counting quantifiers over unary and restricted binary predicates; and in the normal form and linear algebraic methods for Łukasiewicz Infinitely-valued Probabilistic Logic. But the main claim of originality lies in bringing forward the similarities of all those systems.

The following logic systems are studied in detail.

- Probabilistic Logic (PL). It consists of classical propositional logic enhanced with probability assignments over formulas, presented in Sect. 2.
- Counting Quantifiers over a first order fragment containing unary predicates; we show that such a fragment can be extended with binary predicates in restricted contexts without a complexity blow up. The CQU and CQUEL logics are presented in Sect. 3.
- Łukasiewicz Infinitely-valued Probabilistic Logic (LIP), a multi-valued logic for which there exists a well-founded probability theory, presented in Sect. 4.

For each system above, we present language, semantics and decision problem, followed by normal form presentation and satisfiability characterization We also present complexity results and decision algorithms.

It is important to note that throughout this work those logics and their decision problems are presented syntactically, and formulas are linguistic objects, presented as a context-free grammar or some similar, recursive, device. The syntactic vocabulary contains, at the level of terminals, a set of basic (propositional) symbols $\mathscr{P}$, a set of connectives with appropriate arity and punctuation symbols.

2 Probabilistic Logic

Probabilistic logic combines classical propositional inference with classical (discrete) probability theory. The original formulation of such a blend of logic and probability is due to George Boole who, in his seminal work introducing what is now known as Boolean Algebras, already dedicated the two last sections to the problem of combining logic and probability results, stating that

> the object of the theory of probabilities might be thus defined. Given the probabilities of any events, of whatever kind, to find the probability of some other event connected with them.
>
> George Boole [8, Chap. XVI, 4, p. 189]

Deciding if a given set of probabilities is consistent or coherent may be seen as a first step for Boole's "probability extension problem". Indeed, there is certainly more than one way of computing probabilities starting from the establishment of their coherence; see [15] and also the methods presented in this work.

For the purposes of this work, we concentrate on the decision problem of probabilistic logic, the *Probabilistic Satisfiability* problem (PSAT), which consists of an assignment of probabilities to a set of propositional formulas, and its solution consists of a decision on whether this assignment is satisfiable; this formulation is based on a full Boolean Algebra which, due to de Finetti's Dutch Book Theorem (see Proposition 2.5 below), is equivalent to deciding the coherence criterion over a finite Boolean Algebra. The problem has been first proposed by Boole and has since been independently rediscovered several times (see [27, 28] for a historical account) until it was presented to the Computer Science and Artificial Intelligence community by Nilsson [38] and was shown to be an NP-complete problem, even for cases where the corresponding classical satisfiability is known to be in PTIME [24].

Boole's original formulation of the PSAT problem did not consider conditional probabilities, but extensions for them have been developed [27–29, 43]; the latter two works also cover extensions of PSAT with imprecise probabilities. The complexity of the decision problems for conditional probabilities becomes PSPACE-complete if constraints can combine distinct conditional events; otherwise it remains NP-complete [19]. A few tractable fragments of PSAT were presented [1]. In this work, however, we concentrate on PSAT's original formulation, and in this section we follow the developments of [6, 7, 20, 22].

The PSAT problem is formulated in terms of a linear algebraic problem of exponential size. The vast majority of algorithms for PSAT solving in the literature are based on linear programming techniques, such as column generation, enhanced by several types of heuristics [20, 22, 29, 32].

On the other hand, there is a distinct foundational approach to sets of probability assignment to formula known as *coherent probabilities*, which are based on de Finetti's view of probabilities as betting odds [13–15].

In the following we present a few examples in Sect. 2.1, discuss the relationship between PSAT and coherent probabilities in Sect. 2.2 and present an algorithm for deciding PSAT in Sect. 2.3.

2.1 Examples

Consider the following example.

Example 2.1 A doctor is studying a disease D and formulates a hypothesis, according to which there are three genes involved, g_1, g_2 and g_3 such that at least two of which must be present for the disease D to manifest itself. Studies in the population of D-patients shows that each of the three genes is present in 60% of the patients.

The question is whether the doctor's hypothesis is consistent with the data. □

In this example, we see a hypothesis consisting of hard statements (statements with probability 1) being confronted with probabilistic data. The consistency of the joint statement is sees as decision problem of the sort we are dealing with here.

A second example is as follows.

Example 2.2 In an ant colony infestation, three observers have reached different conclusions.

- Observer 1 noticed that at least 75% of the ants had mandibles or could carry pieces of leaves.
- Observer 2 said that at most a third of the ants had mandibles or did not display the ability to carry pieces of leaves.
- Observer 3 stated that at most 15% of the ants had mandibles.

The question is whether these observations are jointly consistent or not. □

We now see how these examples can be formalized.

2.2 Coherent Probabilities and Probabilistic Satisfiability

A *PSAT instance* is a set $\Sigma = \{P(\alpha_i) \bowtie_i p_i | 1 \leq i \leq k\}$, where $\alpha_1, \ldots, \alpha_k$ are classical propositional formulas defined on n logical variables[1] $\mathscr{P} = \{x_1, \ldots, x_n\}$, which are restricted by probability assignments $P(\alpha_i) \bowtie_i p_i$, where $\bowtie_i \in \{=, \leq, \geq\}$ and $1 \leq i \leq k$. It is usually the case that all $\bowtie_i$ are equalities, in which case the PSAT instance can be seen simply as a set of pairs $\{(\alpha_i, p_i) || i = 1, \ldots, k\}$.

There are 2^n possible propositional valuations v over the logical variables, $v : \mathscr{P} \to \{0, 1\}$; each such valuation is truth-functionally extended,[2] as usual, to all formulas, $v : \mathscr{L} \to \{0, 1\}$, and a formula α is *valid* if every valuation satisfies it, noted as $\models \alpha$. Let V be the set of all propositional valuations.

A *probability distribution over propositional valuations*[3] $\pi : V \to [0, 1]$, is a function that maps every valuation to a value in the real interval $[0, 1]$ such that $\sum_{i=1}^{2^n} \pi(v_i) = 1$. The probability distribution π can be uniquely extended over the set of all propositional formulas built from V. This, the probability of a formula α according to distribution π is given by $P_\pi(\alpha) = \sum\{\pi(v_i) | v_i(\alpha) = 1\}$. The following is a straightforward consequence of this definition.

---

[1] In computational logic tradition, variables are also called (syntactical) *atoms*, but to avoid confusion with the algebraic use of 'atom' as the smallest nonzero element of an algebra, we use here instead the term *propositional symbol*, or *(atomic) proposition*.

[2] Thus, valuations can be seen as homomorphisms of the set of formulas into the two element Boolean Algebra $\{0, 1\}$.

[3] While the presentation here stays on the syntactical level, in algebraic terms this notion can be seen as a probability measure over the free boolean algebra, in the sense of [30]. Recall that a measure on A is a function $\tau : A \to [0, 1]$ which is additive for incompatibles and also satisfies $\tau(1) = 1$. When A is finite, as in the case here, every $a \in A$ equals the disjunction of the atoms it dominates, so τ is uniquely determined by its value at the set of (algebraic) atoms of A. For every element $a \in A$ the value of $\tau(a)$ is the sum of the values $\tau(e)$ for all atoms $e \leq a$.

Lemma 2.3 *The probability P_π defined above respects Kolmogorov's basic properties of discrete probability:*

K1 $0 \leq P_\pi(\alpha) \leq 1$

K2 *If* $\models \alpha$ *then* $P_\pi(\alpha) = 1$

K3 *If* $\models \neg(\alpha \wedge \beta)$ *then* $P_\pi(\alpha \vee \beta) = P_\pi(\alpha) + P_\pi(\beta)$

Nilsson [38]'s linear algebraic formulation of PSAT considers a $k \times 2^n$ matrix $A = [a_{ij}]$ such that $a_{ij} = v_j(\alpha_i)$. The *probabilistic satisfiability problem* is to decide if there is a probability vector π of dimension 2^n that obeys the *PSAT restriction*:

$$
\begin{aligned}
A\pi &\bowtie p \\
\sum \pi_i &= 1 \\
\pi &\geq 0
\end{aligned} \tag{1}
$$

where $\bowtie$ is a "vector" of comparison symbols, $\bowtie_i \in \{=, \leq, \geq\}$.

A *PSAT instance* Σ is *satisfiable* iff its associated PSAT restriction (1) has a solution. If π is a solution to (1) we say that π satisfies Σ. The last two conditions of (1) force π to be a probability distribution. Usually the first two conditions of (1) are joined, A is a $(k+1) \times 2^n$ matrix with 1's at its first line, $p_1 = 1$ in vector $p_{(k+1)\times 1}$, so $\bowtie_1$-relation is "=".

Example 2.4 Consider Example 2.1. Let x_i represent that gene i is active in a D-patient. The hypothesis that at least two genes are active in a given D-patient is represented by $\neg(\neg x_i \wedge \neg x_j)$ with 100% certainty for $i \neq j$:

$$P(x_1 \vee x_2) = P(x_1 \vee x_3) = P(x_2 \vee x_3) = 1.$$

The data stating that each gene occurs in 60% of D-patients is given by:

$$P(x_1) = P(x_2) = P(x_3) = 0.6,$$

and the question is if there exists a probability distribution that simultaneously satisfies these 6 probability assignments.

Consider now Example 2.2. Let x_1 mean that an ant has mandibles and x_2 mean that that it can carry pieces of leaves. In this case, we obtain the restrictions Σ:

$$P(x_1 \vee x_2) \geq 0.75 \qquad P(x_1 \vee \neg x_2) \leq 1/3 \qquad P(x_1) \leq 0.15$$

Consider a probability distribution π and all the possible valuations as follows.

π	x_1	x_2	$x_1 \vee x_2$	$x_1 \vee \neg x_2$
0.20	0	0	0	1
0.05	1	0	1	1
0.70	0	1	1	0
0.05	1	1	1	1
1.00	0.10	0.75	0.80	0.30

which jointly satisfies the assignments above, so Example 2.2 is satisfiable. We are going to present an algorithm to compute one such probability distribution if one exists. □

On the other hand, de Finetti's approach aims at defining a "coherent" set of betting odds, or simply a coherent book. Given a map from formulas to real values in [0, 1], $P : \{\alpha_1, \ldots, \alpha_k\} \to [0, 1]$, there is a *Dutch book* against P if there are $\sigma_1, \ldots, \sigma_k \in \mathbb{R}$ such that

$$\sum_{i=1}^{k} \sigma_i (P(\alpha_i) - v(\alpha_i)) < 0 \text{ for all valuations } v.$$

The map is *coherent* if there is no Dutch book against it.

This can be understood as a game between two players, Alice the bookmaker and Bob the bettor, wagging money on the occurrence of α_i. For each i, Alice states her betting odd $P(\alpha_i) = p_i \in [0, 1]$ and Bob chooses a "stake" $\sigma_i \in \mathbb{R}$; Bob pays Alice $\sum_{i=1}^{k} \sigma_i \cdot P(\alpha_i)$ with the promise that Alice will pay back $\sum_{i=1}^{k} \sigma_i \cdot v(\alpha_i)$ if the outcome is possible world (or valuation[4]) v. The chosen stake σ_i is allowed to be negative, in which case Alice pays Bob $|\sigma_i| \cdot P(\alpha_i)$ and gets back $|\sigma_i| \cdot v(\alpha_i)$ if the world turns out to be v. Alice's total balance in the bet is $\sum_{i=1}^{k} \sigma_i (P(\alpha_i) - v(\alpha_i))$. So there is a Dutch book against Alice if the bettor has a choice of stakes such that, for every valuation v, Alice looses money. Thus an assignment is coherent if for every set of stakes a bettor chooses, there is always a possible non-negative outcome. It turns out that coherent maps are precisely those that can be seen as satisfiable PSAT instances.

Proposition 2.5 (de Finetti [13–15]) *Given a map from formulas to real values in* [0, 1], $P : \{\alpha_1, \ldots, \alpha_k\} \to [0, 1]$, *the following are equivalent:*

(a) P *is a coherent book.*
(b) The probability assignment $\Sigma = \{(\alpha_i, P(\alpha_i)) \mid i = 1, \ldots, k\}$ *is a satisfiable PSAT instance.*

As a consequence of Proposition 2.5 and Lemma 2.3, a coherent assignment is one that respects the axioms of probability theory. Furthermore, to decide if an assignment is coherent, we can employ linear algebraic methods that solve (1).

Example 2.6 In Example 2.1, consider a negative stake $\sigma = -1$ for the hypothesis information, and a positive stake of $\sigma = 1$ for the probabilistic data, thus obtaining a total balance of

$$S = -1 \cdot ((1 - v(a \vee b)) + (1 - v(a \vee c)) + (1 - v(b \vee c))) + 1 \quad \cdot ((0.6 - v(a)) + (0.6 - v(b)) + (0.6 - v(c)))$$

[4]The notion of a "world", can be understood via Stone duality, whereby homomorphisms of a boolean algebra A of events into the two element boolean algebra $\{0, 1\}$ are a dual counterpart of A, consisting of all possible evaluations of the events of A into $\{0, 1\}$, and can thus be identified with the set of possible worlds where these events take place.

It turns out that $S < 0$ for all 8 possible worlds v, so this choice of stake constitutes a Dutch Book and the assignment is incoherent and, by Proposition 2.5, it is an unsatisfiable PSAT instance. □

2.3 Algorithms for PSAT Solving

In this presentation, we follow Finger and De Bona [20, 22].

An important result of [24], which is an application of Carathéodory's Theorem [16], guarantees that a solvable PSAT instance has a "small" witness.

Proposition 2.7 *If a PSAT instance $\Sigma = \{P(\alpha_i) = p_i | 1 \leq i \leq k\}$ is satisfiable, then there is a solution. π to the PSAT restrictions (1) such that there at most $k + 1$ elements $\pi_j \geq 0$.* □

Proposition 2.7 implies that the complexity of PSAT is in NP. The special case where all $p_i = 1$ makes classical SAT a special case of PSAT, so PSAT is NP-hard. It follows that PSAT is NP-complete.

A PSAT instance is in *propositional normal form* if it can be partitioned in two sets, $\langle \Gamma, \Psi \rangle$, where $\Gamma = \{P(\alpha_i) = 1 | 1 \leq i \leq m\}$ and $\Psi = \{P(y_i) = p_i | y_i$ is a propositional symbol, $1 \leq i \leq k\}$, with $0 < p_i < 1$. The partition Γ is the SAT part of the normal form, usually represented only as a set of propositional formulas and Ψ is the *propositional probability assignment* part. By adding at most k extra variables, any PSAT instance can be brought to normal form in polynomial time.

Example 2.8 The PSAT instance in Example 2.4 is already in normal form, with $\Gamma = \{x_1 \vee x_2, x_1 \vee x_3, x_2 \vee x_3\}$ and $\Psi = \{P(x_1) = P(x_2) = P(x_3) = 0.6\}$. This indicates that the normal form is a "natural" form in many cases, such as when one wants to confront a theory Γ with the evidence Ψ.

For the formulation of Example 2.2, we add three new variables, y_1, y_2, y_3 and make

$$\begin{aligned} \Gamma &= \left\{ y_1 \rightarrow (x_1 \vee x_2),\ (x_1 \vee \neg x_2) \rightarrow y_2,\ x_1 \rightarrow y_3 \right\} \\ &\equiv \left\{ x_1 \vee x_2 \vee \neg y_1, \neg x_1 \vee y_2, x_2 \vee y_2, \neg x_1 \vee y_3 \right\} \end{aligned}$$

and $\Psi = \{P(y_1) = 0.75, P(y_2) = \frac{1}{3}, P(y_3) = 0.15\}$. □

The algebraic formalization of PSAT (1) has a special interpretation if the formula is in normal form, in which the columns of matrix A are Γ-consistent valuations; a valuation v over $y_1, \ldots, y_k$ is *Γ-consistent* if there is an extension of v over $y_1, \ldots, y_k, x_1, \ldots, x_n$ such that $v(\Gamma) = 1$. This property is the basis for encoding instances of PSAT into those of SAT. However, due to the cubic increase in the number of variables, this method is too inefficient. For details on this form of reduction, see [22].

Instead, we plan to solve (1) without explicitly representing the exponentially large matrix A, using a method called *column generation*. For that, we consider the following linear program:

$$\begin{array}{ll} \min & c' \cdot \pi \\ \text{subject to} & A \cdot \pi = p \\ & \pi \geq 0 \text{ and } \sum \pi_i = 1 \end{array} \tag{2}$$

The cost vector c in (2) is a $\{0, 1\}$-vector such that $c_i = 1$ iff column A^j is Γ-inconsistent. Thus, the column generation process proceeds by generating Γ-consistent columns. The result of this minimization process reaches total cost $c' \cdot \pi = 0$ iff the input instance is satisfiable.

We now describe the column generation process presented in Algorithm 2.1, which solves (2). We start by describing the format of the input data. Condition $\sum \pi_i = 1$ in (2) is usually incorporated in matrix A. By convention, this equation always be the first line of A. Also by convention, vector p is sorted in decreasing order, such that its first position contains a 1, corresponding to the equation $\sum \pi_i = 1$; accordingly, vector p is prefixed with a 1. Let $k = |\Psi|$. This convention allows us to solve the linear program (2) initializing A as an upper triangular matrix T_{up}, which is a $(k + 1) \times (k + 1)$ square matrix where elements on the diagonal and above it are 1 and the remaining ones are 0. As a consequence, the initial probability distribution π is initialized such that $\pi_i = p_i - p_{i+1}$, $1 \leq i \leq k$ and $\pi_{k+1} = p_{k+1}$. The cost c is a $\{0, 1\}$-vector in which $c_j = 1$ iff column A^j is Γ-inconsistent, $1 \leq j \leq k + 1$.

In the column generation process, columns will be added to A, and the vectors for cost c and solution π will be correspondingly extended. As all generated columns at the following steps are Γ-consistent, all cost elements added to c are 0.

Column generation proceeds by steps. At step 0, we start A, c and π as described above (line 1). At each step s, we start by solving the linear program $A_{(s)} \cdot \pi^{(s)} = p$ (line 3); so we suppose there is a linear programming solver available; for an algorithm that does not presuppose a linear solver, see [20]. We require that the solution generated contains the *primal solution* $\pi^{(s)}$ as well as the *dual solution* $z^{(s)}$ [4]; the dual solution is given by $z = c_B \cdot B^{-1}$, where B is the basis of the linear program at step s, that is, a square sub-matrix of A used to compute $\pi^{(s)}$ as the solution of $B\pi^{(s)} = p$, and c_B is the cost of the columns of the basis. These are used in the column generation process (line 4) described below. If column generation fails, then the process cannot decrease current cost and Algorithm 2.1 is terminated with a negative decision. Otherwise, a new column is generated and A and c are expanded. At the end, when the objective function has reached 0, the final values of A and x are returned.

The idea of SAT-based column generation is to map a linear inequality over a set of $\{0, 1\}$-variables into a SAT-formula, using the $O(n)$ method described in Warners [44]. The inequality is provided by the column selection method used by the *Simplex Method* for solving linear programs [4, 39]. Given a linear program in format (2), the *reduced cost* $\bar{c}_y$ of inserting column y from A in a simplex basis is

$$\bar{c}_y = c_y - z' \cdot y \tag{3}$$

where c_y is the cost associated with column y and z is the dual solution of the system $A \cdot \pi = p$ of size $k + 1$. As the generated column y is always Γ-satisfying, $c_y = 0$,

Algorithm 2.1 *PSATViaColGen*(φ)

Input: a normal form PSAT formula $\langle \Gamma, \Psi \rangle$.
Output: a solution (π, A) for (2), if one exists; "No", otherwise.

$A_{(0)} = T_{\text{up}}$; compute cost vector $c^{(0)}$ and $\pi^{(0)}$
for $s = 0;\ c^{(s)\prime} \cdot \pi^{(s)} > 0;\ s{+}{+}$ **do**
  $z^{(s)} = \mathit{DualSolution}(A_{(s)}, p, c^{(s)})$
  $y^{(s)} = \mathit{GenerateColumn}(z^{(s)}, \Gamma)$
  return "No" **if** column generation failed
  $A_{(s+1)} = \mathit{append} - \mathit{column}(A_{(s)}, y^{(s)})$
  $c^{(s+1)} = \mathit{append}(c^{(s)}, 0)$
end for
return $(\pi^{(s)}, A_{(s)})$ such that $A_{(s)} \cdot \pi^{(s)} = p$ and $c^{(s)\prime} \cdot \pi = 0$ // Successful termination

so to ensure a non-increasing value in the objective function we need a non-positive reduced cost, $\bar{c}_y \leq 0$, which leads us to

$$z' \cdot y \geq 0 \tag{4}$$

As y is a $\{0, 1\}$-vector, inequality (4) can be transformed into a SAT-formula; that formula is added to Γ to obtain α, which encodes a solution to (4) that is Γ-satisfying. We then send α to a SAT-solver; if it is unsatisfiable, there is no way to reduce the cost of the linear program's objective function; otherwise, we obtain a satisfying assignment v. The generated column is $v(y)$, the restriction of v the variables in y, which is a solution to (4). A new basis is obtained by substituting $v(y)$ for an appropriate outgoing column. The Simplex Method provides a way of choosing the outgoing column, and guarantees the new basis is a feasible solution to linear program (2) whose cost is smaller than or equal to the previous one.

We have shown how to construct a SAT-based column generation function *GenerateColumn*(z, Γ), provided we are given a (dual) solution for the corresponding linear program.

Theorem 2.9 *Algorithms 2.1 and GenerateColumn provide a decision procedure for the PSAT problem.*

Proof The correctness of Algorithm 2.1 is a direct consequence from the fact that $\langle \Gamma, \Psi \rangle$ is satisfiable iff the linear program (2) reaches a minimum at 0. As column generation only fails when it is impossible to decrease the cost function, this process either fails or brings the cost to 0, which is the only way Algorithm 2.1 terminates with a solution. □

Note that the proof above guarantees termination, but even if it uses a polynomial-time linear solver, no polynomial bound is provided for the number of steps, which can in principle be $O(2^k)$. Several implementations using the simplex method, employing various column generation strategies, are described in Finger and De Bona [22], which also describe important empirical properties of those implementation.

3 Counting Quantifiers over Unary Predicates

Counting quantifiers are quantitative constraints which may superficially look different from probabilistic reasoning. However, here we demonstrate that there are striking similarities between these two forms of reasoning allowing us to lump them together under the heading of Quantitative Logic Reasoning.

The need to combine deductive reasoning with counting and cardinality capabilities in a principled way has been long recognized, but the complexity of this task has precluded its development. However, by generating a fragment of counting quantifiers inspired by the PSAT formulation, we are able to present a useful deductive system with counting that is "only" NP-complete and which allows for reasonably efficient, deterministic algorithms.

The basic approach for adding counting capabilities extends first-order logic with some form of generalized quantifiers [35], and we employ here a Lindstrom-type of quantifier [33] that can express the counting notions of "there are at least/most n elements with property P". Counting is first-order expressible, but it requires a first-order encoding using at least as many symbols as the counts one aims to express. On the other hand, the number of symbols employed by counting quantifiers is only proportional to the number of digits of the counts expressed. Hence expressing counting in first-order logic results in formulas whose size is exponentially larger than those obtained by using counting quantifiers.

The satisfiability of a logic with counting quantifiers, but limited to a two-variable fragment with at most binary predicates, is decidable [25, 26] with an EXPTIME-hard lower bound [2] and a NEXPTIME [40] upper-bound[5]; recent studies on the complexity of specific counting problems are found in Martin et al. [34]; Bulatov and Hedayaty [10]. Focusing on a one-variable fragment containing counting quantifiers over unary predicates only, the decision problem becomes NP-complete, even when restricted only to a fragment called Syllogistic Logic, but the decision algorithm used to show that is inherently non-deterministic [41].

In a previous work, we presented an expressive fragment of first-order logic with counting quantifiers over unary predicates called CQU [21], which was developed applying techniques similar to those used in the PSAT case. Here we extend the work on CQU by introducing CQUEL, for which the satisfiability problem remains NP-complete even as it partially allows the presence of binary predicates. We start by presenting CQU, extend it to CQUEL, and then develop decision algorithms for it, We start with a general example.

Example 3.1 Consider the following group of people with several ages

(a) At most 15 grandparents are married or happy.
(b) At least 10 parents are not happy.
(c) At most 7 parents are not married.
(d) All grandparents are parents.

---

[5]Note that the fragment mentioned here has the finite model property [41].

(e) At most 7 grandparents are unmarried and unhappy.
(f) At least 8 grandparents are unmarried and unhappy.

We would like a way to determine that (a)–(d) are satisfiable. We would also like to have a method that allows us to infer (e) from those statements; equivalently, as (f) can be seen as the negation of (e), determine that (a)–(d) and (f) are jointly unsatisfiable. These possibilities are all covered by the CQU formalism. Moreover, suppose we are given a list of parent-children pairs (i.e. a binary relation), and define a parent as someone who has a child and, likewise, a grandparent as someone who has a grandchild. To deal with this more general formulation, one needs a more expressive formalism such as CQUEL. □

In the following, we present the Semantics of CQU (Sect. 3.1) and its extension CQUEL (Sect. 3.2). Then we present an algebraic formulation of the CQUEL-SAT problem (Sect. 3.3) which is used as a basis for the algorithms that solve it (Sect. 3.4).

3.1 Semantics and Satisfiability of CQU

We now present formally a function-free one-variable first-order fragment over a signature containing only unary predicates and constants, extended with explicit counting quantifiers $\exists_{\leq n}$ (at most n) and $\exists_{\geq n}$ (at least n), where $n \in \mathbb{N}$ is a non-negative integer. The semantics is tarskian, with models of arbitrarily large cardinality.

The fragment contains two types of sentences over a countable set of variables V. Let $\psi(x)$ be a Boolean combination of unary predicates $p(x)$, $q(x)$, etc. A *counting sentence* has the form $\exists_{\leq n} x\ \psi(x)$ or $\exists_{\geq n} x\ \psi(x)$. A *universal sentence* has the form $\forall x \psi(x)$. A formula φ over the fragment of counting quantifiers over unary predicates (CQU), is a conjunction of any finite number of counting sentences $\mathscr{Q}$ and universal sentences $\mathscr{U}$, $\varphi = \langle \mathscr{Q}, \mathscr{U} \rangle$. Note that the universal and counting sentences involve only one-variable and only unary predicates; when we introduce the CQUEL fragment below a two-variable fragment will be involved, with restricted use of binary predicates.[6]

For the semantics, let the *domain* D be a non-empty set. Let a *term* be a constant or a variable. Consider an *interpretation* $\mathscr{I}$; when applied to a term t, $\mathscr{I}(t) \in D$ and when applied to a unary predicate p, $\mathscr{I}(p) \subseteq D$; $\mathscr{I}_{|x}$ represents an interpretation that is identical to $\mathscr{I}$, except possibly for the interpretation of x. Let φ be a CQU-formula; by $\mathscr{I} \models \varphi$ we mean that φ is satisfiable over D with interpretation $\mathscr{I}$, defined as

$$\begin{array}{ll} D, \mathscr{I} \models p(t) & \textit{iff } \mathscr{I}(t) \in \mathscr{I}(p) \\ D, \mathscr{I} \models \neg\psi & \textit{iff } D, \mathscr{I} \not\models \psi \\ D, \mathscr{I} \models \psi \wedge \rho & \textit{iff } D, \mathscr{I} \models \psi \text{ and } D, \mathscr{I} \models \rho \\ D, \mathscr{I} \models \exists_{\leq n} x\ \psi & \textit{iff } \left|\{\mathscr{I}_{|x}(x) \in D | D, \mathscr{I}_{|x} \models \psi\}\right| \leq n \\ D, \mathscr{I} \models \exists_{\geq n} x\ \psi & \textit{iff } \left|\{\mathscr{I}_{|x}(x) \in D | D, \mathscr{I}_{|x} \models \psi\}\right| \geq n \end{array}$$

[6]First-order one- and two-variable fragments are decidable, but the coding of counting quantifiers employs several new variables, so decidability is not immediate; see Proposition 3.2.

The usual definitions apply to other Boolean connectives. Note that the negation of counting sentences can be expressed within the CQU fragments, namely $\neg\exists_{\leq n}x\ \psi \equiv \exists_{\geq n+1}x\ \psi$ and $\neg\exists_{\geq n+1}x\ \psi \equiv \exists_{\leq n}x\ \pi$. The first-order existential quantifier is expressed as $\exists x\psi \equiv \exists_{\geq 1}x\ \psi$ and the universal quantifier as $\forall x\psi \equiv \exists_{\leq 0}x\ \neg\psi$. The *exact counting quantifier* is defined as $\exists_{=n}x\ \psi \equiv \exists_{\leq n}x\ \psi \wedge \exists_{\geq n}x\ \psi$.

If there are D and $\mathscr{I}$ such that $D, \mathscr{I} \models \varphi$, then φ is a *satisfiable* formula; otherwise it is unsatisfiable. A formula φ *entails* ψ ($\varphi \models \psi$) iff every pair $(D, \mathscr{I})$ that satisfies the former also satisfies the latter. φ is *equivalent* to ψ ($\varphi \equiv \psi$) iff they are satisfied by the same pairs $(D, \mathscr{I})$. The problem CQU-SAT consists of deciding whether a formula is satisfiable.

If we remove the restriction to conjunctions of universal and counting sentences, we obtain the fragment called $\mathscr{C}^1$, studied in Pratt-Hartmann [41]. Unlike CQU, $\mathscr{C}^1$ allows for disjunctions between quantified formulas, such as $\exists_{\geq 7}x\ \psi \vee \exists_{\leq 9}y\ \rho$. As the $\mathscr{C}^1$ fragment has the finite model property and contains CQU, we obtain the following.

Proposition 3.2 (Pratt-Hartmann [41]) *Every satisfiable CQU formula is satisfiable over a finite domain. Moreover, CQU-SAT is strongly NP-complete.* □

Strong NP-completeness means that when n in $\exists_{\leq n}$, $\exists_{\geq n}$ is given in unary notation, the decision remains NP-complete. As with probabilistic logic, we propose a normal form for formulas in the CQU fragment. Existence of such normal form for CQU will allow us to extend the method to CQUEL.

Definition 3.3 Let $\mathscr{U}$ be a finite set of universal sentences and let $\mathscr{Q}$ be a finite set of quantified unary predicates of the form $\exists_{\leq n}x\ p(x)$ or $\exists_{\geq n}x\ p(x)$, where p is a unary atomic predicate. A *normal form* CQU formula $\varphi = \langle \mathscr{Q}, \mathscr{U} \rangle$ is the conjunction of formulas in $\mathscr{Q} \cup \mathscr{U}$.

In the following we use $\bowtie$ to refer to $\leq$ or $\geq$, so the CQU normal form is characterized by counting quantifier sentences of the form $\exists_{\bowtie n}x\ p(x)$. By adding a small number of extra predicates, any CQU formula can be brought to normal form.

Lemma 3.4 *For every CQU formula φ there exists a normal form formula φ' such that φ is a satisfiable iff φ' is; the normal form φ' can be built from φ in polynomial time.*

Proof Consider $\varphi = \langle \mathscr{Q}, \mathscr{U} \rangle$. We build $\varphi' = \langle \mathscr{Q}', \mathscr{U}' \rangle$ starting with $\mathscr{Q}' = \emptyset$ and $\mathscr{U}' = \mathscr{U}$. Then, for every quantified formula $\exists_{\bowtie n}x\ \psi$, if ψ is an unary predicate, just add $\exists_{\bowtie n}x\ \psi$ to $\mathscr{Q}'$; otherwise, create a new unary predicate p_{new} and add $\forall x(p_{\text{new}}(x) \leftrightarrow \psi)$ to $\mathscr{U}'$ and $\exists_{\bowtie n}x\ p_{\text{new}}(x)$ to $\mathscr{Q}'$; at every step φ' is in normal form, and at its end, by construction, φ' is satisfiable iff φ is. □

Example 3.5 Consider Example 3.1, which can be formalized as follows:

(a) $\exists_{\leq 15}x\ (g(x) \wedge (m(x) \vee h(x)))$
(b) $\exists_{\geq 10}x\ (g(x) \wedge \neg h(x))$

(c) $\exists_{\leq 7} x\,(p(x) \wedge \neg m(x))$
(d) $\forall x (g(x) \rightarrow p(x))$
(e) $\exists_{\leq 7} x\,(g(x) \wedge \neg m(x) \wedge \neg h(x))$
(f) $\exists_{\geq 8} x\,(g(x) \wedge \neg m(x) \wedge \neg h(x))$

Clearly, (e) is the negation of (f); we use only the latter. To bring counting formulas to normal form, introduce the predicates, q_1, q_2, q_3, q_4. Let $\mathscr{U} = \{\forall x (q_1(x) \leftrightarrow (g(x) \wedge (m(x) \vee h(x))), \forall x (q_2(x) \leftrightarrow (g(x) \wedge \neg h(x)), \forall x (q_3(x) \leftrightarrow (p(x) \wedge \neg m(x)), \forall x (g(x) \rightarrow p(x)), \forall x (q_4(x) \leftrightarrow (g(x) \wedge \neg m(x) \wedge \neg h(x))\}$, so that we can have counting quantification over unary predicates only; let $\mathscr{Q} = \{\exists_{\leq 15} x\, q_1(x), \exists_{\geq 10} x\, q_2(x), \exists_{\leq 7} x\, q_3(x)\}$, such that we expect $\langle \mathscr{Q}, \mathscr{U} \rangle$ to be satisfiable and $\langle \mathscr{Q} \cup \{\exists_{\geq 8} x\, q_4(x)\}, \mathscr{U} \rangle$ to be unsatisfiable. □

It is important to note that the satisfiability problem for a set of CQU universal formulas is an NP-complete problem, for it can be reduced to a propositional problem.

In fact, consider a normal form $\varphi = \langle \mathscr{Q}, \mathscr{U} \rangle$. Consider the $k = |\mathscr{Q}|$ unary predicates occurring in $\mathscr{Q}$, $p_1(x), \ldots, p_k(x)$; then there are 2^k *elementary terms* of the form $e(x) = \lambda_1(x) \wedge \ldots \wedge \lambda_k(x)$, where each $\lambda_i(x)$ is either $p_i(x)$ or $\neg p_i(x)$; an elementary term $e(x)$ is called *susceptible* if it is consistent with the universal sentences, that is, the set $\{\exists x e(x)\} \cup \mathscr{U}$ has a model.

Semantically, each elementary term is interpreted as a *domain elementary subset* $E \subseteq D$, $E = L_1 \cap \ldots \cap L_k$, where each L_i is either the interpretation of p_i or its complement with respect to the domain D. In any interpretation that satisfies $\varphi = \langle \mathscr{Q}, \mathscr{U} \rangle$, only susceptible elementary terms may be interpreted as non-empty elementary subsets, otherwise the interpretation falsifies $\mathscr{U}$.

Lemma 3.6 *The problem of determining if there exists an elementary domain subset over the unary predicates in $\mathscr{Q}$ that is susceptible with a set of CQU universal formulas $\mathscr{U}$ is NP-complete.*

Proof Transform the set $\mathscr{U}$ into a propositional formula, by deleting the external $\forall x$ quantifiers and considering each unary predicate $p(x)$ as a propositional symbol p. Then determining the existence of a satisfying valuation is an NP-complete problem. If there is such a valuation, we obtain a susceptible element by considering a satisfying valuation v restricted to the proposition corresponding to the unary predicates in $\mathscr{Q}$. In that case, we consider a singleton domain $D = \{d\}$ and an interpretation $\mathscr{I}$ such that $d \in \mathscr{I}(p)$ iff $v(p) = 1$. □

We now expand these results of Finger and Bona [21] to include universal quantification over binary relations. The aim is to maintain the decision problem in NP.

3.2 Expanding CQU into CQUEL

Previous results involving counting quantifiers and binary predicates brought the complexity of the satisfiability problem into EXPTIME [2]. However those methods allowed for counting quantification over sentences involving binary predicates. The idea here is to maintain counting quantification over unary predicates in $\mathscr{Q}$, but to expand the set of allowed sentences in $\mathscr{U}$ so as to maintain the complexity of $\mathscr{U}$-satisfiability in NP.

Our idea is to expand $\mathscr{U}$ to allow for sentences corresponding to the first-order translation of statements from description logic DL Light [11], thus lightly expanding CQU into CQUEL. The satisfiability problem for DL Light is tractable, and we show here that adding those formulas to $\mathscr{U}$ leaves the satisfiability complexity in NP. There is at least another well known tractable description logic, $\mathscr{E}\mathscr{L}^{++}$, which is however maximal with respect to tractability, in the sense that extending its language with the expressivity of CQU universal sentences brings the complexity to EXPTIME-complete [3].

The first-order signature now contains a finite set $\mathscr{P}$ of unary predicates, a finite set $\mathscr{R}$ of binary relations and a finite set $\mathscr{C}$ of constants. The set of *basic concepts* is the smallest set of unary expressions such that:

- every unary predicate $p \in \mathscr{P}$ is a basic concept;
- if $r \in \mathscr{R}$, then $\exists y\, r(x, y)$ and $\exists y\, r(y, x)$ are basic concepts.

Basic concepts form *concepts* in the following way.

- every basic concept $B(x)$ is a concept;
- if $B(x)$ is a basic concept, then $\neg B(x)$ is a concept;
- if $C_1(x)$ and $C_2(x)$ are concepts, so is $C_1(x) \wedge C_2(x)$.

A set $\mathscr{E}$ of *extended light (EL) constraints* is a finite set of universal formulas of the following form.

(a) *Inclusion Assertion (IA)*: $\forall x(B(x) \to C(x))$, where $B(x)$ is a basic concept and $C(x)$ is a concept;

(b) *Functionality Assertion (FA)*: $\mathsf{Funct}(r)$ and $\mathsf{Funct}(r^-)$, for $r \in \mathscr{R}$. The semantics of $\mathsf{Funct}(r)$ states that if $(d, d'), (d, d'') \in \mathscr{I}(r)$, then $d' = d''$; similarly, the semantics of $\mathsf{Funct}(r^-)$ states that if $(d', d), (d'', d) \in \mathscr{I}(r)$, then $d' = d''$.

(c) *Data:* $p(a), r(a, b)$ for $a, b \in \mathscr{C}$, $p \in \mathscr{P}$ and $r \in \mathscr{R}$.

Note that EL constraints, except FAs, belong to a two-variable first-order fragment; FAs require the use of three variables, however in a very limited way. It turns out that a the consistency of a set of EL constraints is not only decidable, but even tractable.

A set of constraints $\mathscr{E}$ is *negative inclusion (NI) closed* if for every IA $A = \forall x(B_1(x) \to \neg B_2(x))$ above such that $\mathscr{E} \models A$, then $A \in \mathscr{E}$. The *NI-closure* of a set of EL constraints $\mathscr{E}$, $\overline{\mathscr{E}}$, is the smallest NI-closed set of constraints that contains $\mathscr{E}$. The tractability of the satisfiability of a set of EL constraints follows from the following result.

Proposition 3.7 (Calvanese et al. [11]) *The NI-closure $\overline{\mathscr{E}}$ can be computed in polynomial time on the number of EL constraints in $\mathscr{E}$.*

The proof of Proposition 3.7 involves defining a normal form for constraints, then showing that computing the NI-closure of a set of constraints can be reduced in linear time to computing the NI-closure of a normalized set of constraints. Then a set of constraint inference rules is proposed and it is shown that: (a) each rule application can be decided in polynomial time and the maximum number of rule applications is polynomial in $|\mathscr{E}|$; (b) each expansion rule inserts only an inferable constraint, from which soundness follows; (c) the inconsistency of $\overline{\mathscr{E}}$ can be obtained by a simple pattern search, which can be decided also in polynomial time. By composing all steps, we have a satisfiability check performed in polynomial time. Details in Calvanese et al. [11].

We now define a CQUEL formula $\varphi = \langle \mathscr{Q}, \mathscr{U}, \mathscr{E} \rangle$ as a conjunction of the counting sentences in $\mathscr{Q}$, the universal CQU sentences in $\mathscr{U}$ and the EL constraints in $\mathscr{E}$. The semantics of a CQUEL formula φ, $D, \mathscr{I} \models \varphi$, is exactly as before, with the addition of binary predicates as in regular first-order logic.

Example 3.8 Consider Example 3.5, which we now extend with a binary relation *parentOf*(x, y) representing the fact that x is a parent of y. Then we add the following set of EL constraints, stating that a parent is a parent-of someone.

$$\mathscr{E} = \Big\{ \forall x \Big(\; p(x) \leftrightarrow \exists y \, \mathit{parentOf}(x, y) \Big) \Big\}$$

The previous result on the existence of normal forms applies to CQUEL too, in which counting quantification is applied only to unary atomic predicates.

Lemma 3.9 *For every CQUEL formula φ there exists a polynomial-time computable normal-form formula $\varphi' = \langle \mathscr{Q}, \mathscr{U}, \mathscr{E} \rangle$ such that φ is a satisfiable iff φ' is, where $\mathscr{Q}$ contains only counting sentences over unary predicates.*

Proof Following Lemma 3.4, the counting and universal sentences in φ are brought to normal form, the EL constraints of φ are also brought to normal form and added as $\mathscr{E}$ to φ'. Clearly, this can be done in polynomial time and satisfiability is preserved by Lemma 3.4. □

The following is a step into showing that the complexity of CQUEL is no greater than that of CQU.

Lemma 3.10 *The problem of deciding if a a set of formulas $\mathscr{U} \cup \mathscr{E}$ is consistent is NP-complete, where $\mathscr{U}$ is a set of CQU universal formulas and $\mathscr{E}$ is a set of EL constraints.*

Proof Extend $\mathscr{E}$ into $\mathscr{E}'$, such that, for every existential constraint there is a new unary predicate equivalent to it; clearly this extension can be done in linear time and the input set $\mathscr{U} \cup \mathscr{E}$ is satisfiable iff $\mathscr{U} \cup \mathscr{E}'$ is. Compute the NI-closure $\overline{\mathscr{E}'}$ in polynomial time, by Proposition 3.7. The inconsistency of $\overline{\mathscr{E}'}$ can be decided in polynomial time, and if it is inconsistent so is the input set.

If $\overline{\mathscr{E}'}$ is consistent, we construct $\mathscr{U}'$ by extending $\mathscr{U}$ with the IAs in $\overline{\mathscr{E}'}$ which contain only unary predicates; that is, there are no binary formulas in $\mathscr{U}'$. By Lemma 3.6, the coherence detection of $\mathscr{U}'$, and thus its satisfiability, is an NP-complete problem.

Clearly, if $\mathscr{U}'$ is unsatisfiable, so is the input set. If it is satisfiable, then we claim that $\mathscr{U} \cup \mathscr{E}$ is also satisfiable. In fact, if $\mathscr{U}'$ is satisfiable, there is a model for it and by the proof of Lemma 3.6 there is a model $(D, \mathscr{I})$ satisfying $\mathscr{U}'$. We extend this model in the following way. Create a set S of facts, initially empty. For each fact $p(a)$ or $r(a, b)$ in $\overline{\mathscr{E}'}$, add this fact to S and start the update propagation process.

The update propagation process consists of the following. If there is a constant a in S for which $\mathscr{I}(a)$ is not defined, create a new element and update $\mathscr{I}$ with the constant and predicate interpretation; then propagate this update up the IA chain with constraints $\forall x\,(B(x) \to C(x)) \in \overline{\mathscr{E}'}$. We have to deal with four cases. If $C(a) = p_i(a)$, add $p_i(a)$ to S and propagate this change. If $C(a) = \neg p_i(a)$, do not update $\mathscr{I}$; due to the consistency of $\overline{\mathscr{E}'}$, we know that $p_i(a)$ will never be added to S. If $C(a) = \exists y r_j(a, y)$ and there is no pair $(\mathscr{I}(a), d) \in I(r_j)$ skolemize $\exists y r_j(a, y)$ by taking a fresh constant a', adding $r_j(a, a')$ to S, and propagate. If $C(a) = \neg\exists y r_j(a, y)$, do not update $\mathscr{I}$; again due to the consistency of $\overline{\mathscr{E}'}$, we know that $r_j(a, b)$ will never be added to S. Similarly, no violation of a functionality assertion can occur, due to the consistency of $\overline{\mathscr{E}'}$.

As the number of possible updates is finite, this update propagation process finishes in a finite number of steps, and we end up with an updated model $(D, \mathscr{I})$ that satisfies both $\overline{\mathscr{E}'}$ and $\mathscr{U}'$, and thus satisfies the input set, as desired. □

The proof of Lemma 3.10 gives us an Algorithm 3.1 to determine the joint satisfiability of $\mathscr{U} \cup \mathscr{E}$. Line 8 employs a SAT-solver, such as [5, 17].

The basic idea of Algorithm 3.1 is to compose a formula to submit it to a SAT solver. For that, the NI-closure of the input set of EL constraints is first computed. If that already shows the problem is unsatisfiable, return. Otherwise construct a propositional formula based on the "propositional core" of $\mathscr{U}$ and $\mathscr{E}$. The final solution is obtained from applying a SAT-solver to this propositional formula.

For the rest of this work we always assume that formulas are in normal form. In the following, we look at CQUEL satisfiability in terms of integer linear algebra.

Algorithm 3.1 JointSAT($\mathscr{U}$, $\mathscr{E}$)

Input: A set of CQU universal sentences $\mathscr{U}$ and a set of EL constraints $\mathscr{E}$.
Output: If satisfiable, return a valuation representing a susceptible term; or "No", if unsatisfiable.

```
1: Extend ℰ into ℰ′, adding for every existential constraint a new unary predicate equivalent to it;
2: Compute the NI-closure ℰ′‾;
3: if ℰ′‾ is inconsistent then
4:     return "No";
5: end if
6: Extend 𝒰 into 𝒰′, adding the IAs in ℰ′‾ which contain only unary predicates;
7: Transform 𝒰′ into a propositional formula α, removing the quantifiers and variables;
8: Apply SAT solver to α;
9: if α is satisfiable then
10:     return satisfying valuation;
11: else
12:     return "No"
13: end if
```

3.3 Algebraic Formulation of CQUEL-SAT

Consider a normal form CQUEL formula $\varphi = \langle \mathscr{Q}, \mathscr{U}, \mathscr{E} \rangle$ whose satisfiability we want to determine. Consider the $k = |\mathscr{Q}|$ unary predicates occurring in $\mathscr{Q}$, $p_1(x), \ldots, p_k(x)$; as in the CQU case, there are 2^k *elementary terms* of the form $e(x) = \lambda_1(x) \wedge \ldots \wedge \lambda_k(x)$, where each $\lambda_i(x)$ is either $p_i(x)$ or $\neg p_i(x)$; an elementary term $e(x)$ is called *susceptible* if it is consistent with $\mathscr{U} \cup \mathscr{E}$, that is, the set $\{\exists x e(x)\} \cup \mathscr{U} \cup \mathscr{E}$ has a model. Only susceptible elementary terms may be interpreted as non-empty elementary subsets, otherwise the interpretation falsifies $\mathscr{U} \cup \mathscr{E}$.

An integer linear algebraic presentation of CQUEL-SAT is based on encoding each elementary term $e(x)$ as a $\{0, 1\}$-vector e of size k, in which $e_i = 1$ if λ_i is p_i in $e(x)$ and $e_i = 0$ otherwise. We consider only the set of k_m susceptible elementary terms, $0 \leq k_m \leq 2^k$. Let A be a $k \times k_m$ $\{0, 1\}$-matrix, where each column encodes a susceptible elementary term; note that the ith line corresponds to the ith counting quantifier expression in $\mathscr{Q}$. Let the ith element in $\mathscr{Q}$ be $\exists_{\bowtie_i n_i} x\ p_i(x)$, $\bowtie_i \in \{\leq, \geq\}$; let b be a $k \times 1$ integer vector, such that $b_i = n_i$, and let x be a $k_m \times 1$ vector of integer variables. Then the potentially exponentially large integer linear system that corresponds to the CQUEL-SAT problem $\varphi = \langle \mathscr{Q}, \mathscr{U}, \mathscr{E} \rangle$ is:

$$\begin{aligned} Ax &\bowtie b \\ x &\geq 0 \\ x_j &\quad \text{integer} \end{aligned} \tag{5}$$

Lemma 3.11 *A normal form $\varphi = \langle \mathscr{Q}, \mathscr{U}, \mathscr{E} \rangle$ is CQUEL satisfiable iff its corresponding system given by* (5) *has a solution.*

Proof (⇒) If φ has an interpretation, let x_j be the number of elements in the elementary subset corresponding to the jth susceptible elementary term; clearly x_j is a non-negative integer. As all elements in $\mathscr{Q}$ are satisfied, all inequalities in $Ax \bowtie b$ are satisfied.

(⇐) If system (5) has a solution, we construct a finite interpretation by inserting x_j elements in each subset corresponding to a susceptible elementary term. We can then compute an interpretation for all predicates in $\mathscr{Q}$, and as all inequalities in (5) are satisfied, so is $\mathscr{Q}$; furthermore, as only susceptible elementary terms have non-zero elements, $\mathscr{U} \cup \mathscr{E}$ is also satisfied. □

To determine if an elementary term is susceptible, apply Algorithm 3.1, and consider the part of the returned valuation corresponding to the predicates in $\mathscr{Q}$, as illustrated in the following example.

Example 3.12 Consider the normal form formula $\varphi = \langle \mathscr{Q}, \mathscr{U}, \mathscr{E} \rangle$ presented in Examples 3.5 and 3.8. The linear algebraic rendering of the problem shows it is CQUEL satisfiable:

$$
\begin{array}{ll}
\mathbf{q_1} & \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} 10 \\ 0 \\ 0 \end{bmatrix} \begin{matrix} \leq 15 \\ \geq 10 \\ \leq 7 \end{matrix} \\
\mathbf{q_2} & \\
\mathbf{q_3} & \\
\hdashline
\mathbf{g} & 0\ 1\ 0 \\
\mathbf{p} & 1\ 0\ 1 \\
\mathbf{m} & 1\ 0\ 0 \\
\mathbf{h} & 0\ 1\ 1
\end{array}
$$

Then first three columns {0, 1}-columns of size 7 are valuations over all unary predicates satisfying $\mathscr{U} \cup \mathscr{E}$; each valuation represents an elementary domain over predicates which are assigned 1 and the complement of the 0-assigned predicates. Each line corresponds to a predicate, indicated on the left. The top three lines contain the quantified restrictions in $\mathscr{Q}$ and the matrix-vector product satisfies the counting inequalities; the last four lines correspond to the predicates whose count are not quantified in $\mathscr{Q}$. The three . This solution implies that the first four conditions of Examples 3.1 and 3.5 are satisfiable.

However, to show that adding the last condition leads to an unsatisfiable set of sentences, we would have to exhaustively consider the 2^4 valuations over predicates $q_1, \ldots, q_4$ and show that that exponentially large system cannot satisfy the 4 inequalities. □

The exponential size of the proof search alluded by Example 3.12 can be avoided if there is a guarantee that all satisfiable CQUEL formulas have polynomial-sized models. In the case of Probabilistic Satisfiability (PSAT), which does not have the restriction on integral solution, the existence of polynomial-size models is guaranteed by Caratheodory's Theorem [16]. In the discrete case, we have the following analogue, which provides a polynomial-sized bound for models of satisfiable CQUEL-SAT.

Proposition 3.13 (Pratt-Hartmann [41], Eén and Sörensson [18]) *Consider a system of inequalities of the format* (5) *that has a positive integral solution. Then it has a positive integral solution with at most* $\left(\frac{5}{2}k \log k + 1\right)$ *non-zero entries.*

As presented in Algorithm 3.1, the satisfiability of $\mathscr{U} \cup \mathscr{E}$ can be represented by a {0, 1}-valuation representing a susceptible term over its predicates. Let {0, 1}-matrix A be as in (5); A's jth column A^j is *satisfying* if it represents the bits of a valuation returned by Algorithm 3.1 on input $\mathscr{U} \cup \mathscr{E}$ restricted to $\mathscr{Q}$'s. Lemma 3.11 and Proposition 3.13 yield the following.

Lemma 3.14 *Consider a normal form CQUEL-SAT instance* $\varphi = \langle \mathscr{Q}, \mathscr{U}\mathscr{E} \rangle$. *Then* φ *is satisfiable iff there exists a solvable system of inequalities of the form*

$$A_{k \times k_m} \cdot x_{k_m \times 1} \bowtie b_{k \times 1} \tag{6}$$

where $k_m \leq \left\lceil \frac{5}{2}(k \log k + 1) \right\rceil$, *A is a* {0, 1}*-matrix whose columns satisfy* $\mathscr{U} \cup \mathscr{E}$. □

This serves as a basis for effective algorithms for CQUEL-SAT.

3.4 A CQUEL-SAT Solver Based on Integer Linear Programming

The polynomial-size format of solutions given by Lemma 3.14 provides a way to reduce a CQUEL-SAT to SAT; that is, an instance $\varphi = \langle \mathcal{Q}, \mathcal{U} \rangle$ of a CQUEL-SAT decision problem is polynomially translated to an instance of SAT by encoding the set of inequalities in (6) such that the CQUEL-SAT is satisfiable iff its SAT translation is. This approach is described in Finger and Bona [21], but the high number of variables in the translated SAT formulas, which is $O(k^3 \log k)$, makes this approach impractical in the critical areas of hard problems. So a different approach, based on integer linear programming (ILP) and the branch-and-bound algorithm will be pursued.

The algebraic formulation of CQUEL-SAT on input $\varphi = \langle \mathcal{Q}, \mathcal{U}, \mathcal{E} \rangle$ given by (5) is apparently suited for Integer Linear Programming (ILP), finding a solution to $Ax \bowtie b$, where $x_j \in \mathbb{N}$. However, there are two important facts in (5) that have to be addressed, namely

- Matrix A may be exponentially large.
- As a consequence, we do not represent matrix A explicitly; instead we deal with it partially and implicitly.

In fact, A's columns consists of $\{0, 1\}$-valuations representing susceptible terms satisfying $\mathcal{U} \cup \mathcal{E}$, which are costly to compute and there may be exponentially many, e.g. when $\mathcal{U} = \mathcal{E} = \emptyset$. To avoid these problems, we propose to solve the ILP problem via a simplified version of the *branch-and-bound* algorithm [42], which solves relaxed (continuous) linear programs. As in the case of PSAT, we generate A's column as needed, in the process of *column generation* [31] which takes place at each relaxed problem created by the branch-and-bound approach. For an ILP of the form (5), it is not necessary to search for an optimal integer solution, one only needs to find a feasible one or show none exists.

The branch-and-bound method traverses an implicit search tree of relaxed problems. The top level of this search method is shown in Algorithm 3.2. It starts in the root of the search tree with a unary set of problems containing the input CQUEL formula, and it loops until either a feasible integer solution to the corresponding linear algebraic problem given by (6) is found or the set of problems becomes empty, in which case an unsatisfiability decision is reached. In the main loop (lines 3–15), a problem is heuristically selected from the set of problems (line 4), and its *relaxed version* is solved, which consists of the same problem without the restriction of integral solutions. The heuristics implemented orders the problems according to the relaxed solutions to its parent in the tree, giving preference to solutions with the largest number of integer components.

If the relaxed problem has no solution, it is just removed from the set, which corresponds to closing a branch in the search tree, and the next iteration starts searching at an open branch. If there is an integer solutio, the problem is satisfiable and the loop ends. Otherwise, a solution with at least one non-integral element exists. A sec-

Algorithm 3.2 CQUELBranchAndBound(φ)

Input: A normal form CQUEL formula $\varphi = \langle \mathscr{Q}, \mathscr{U}, \mathscr{E} \rangle$.
Output: A solution satisfying (6); or "No", if unsatisfiable.

```
1: CQUELSet = {φ}
2: SAT = false
3: while not SAT and CQUELSet is not empty do
4:    CQUELProblem = RemoveHeuristically(CQUELSet)
5:    solution = SolveRelaxedViaColGen(CQUELProblem)
6:    if no solution found then
7:       continue
8:    else if integral solution then
9:       SAT = true
10:   else
11:      var = choseBranchVar(solution)
12:      newCQUELs = boundedProblems(CQUELProblem, var)
13:      CQUELSet = CQUELSet ∪ newCQUELs
14:   end if
15: end while
16: if SAT then
17:    return solution
18: else
19:    return "No"
20: end if
```

ond heuristics is used to find a variable x_i with a non-integral solution z_i on which to branch (line 11), creating two new branches on the search tree. This heuristics chooses x_{i^*} for which the non-integral z_{i^*} is closer to either $\lfloor z_{i^*} \rfloor$ or $\lceil z_{i^*} \rceil$.

The branching generates two new bounded problems $\varphi' = \langle \mathscr{Q}', \mathscr{U}', \mathscr{E} \rangle$, $\varphi'' = \langle \mathscr{Q}'', \mathscr{U}'', \mathscr{E} \rangle$ (line 12), with the creation of a new unary predicate p_{new}. Note that the set of constraints $\mathscr{E}$ is never changed. We make $\mathscr{U}' = \mathscr{U}'' = \mathscr{U} \cup \{\forall x(p_{\text{new}}(x) \leftrightarrow e_{i^*}(x))\}$, where $e_{i^*}(x)$ is the elementary term corresponding to column i^* and $\mathscr{Q}' = \mathscr{Q} \cup \{\exists_{\leq \lfloor z_{i^*} \rfloor} x\ p_{\text{new}}(x)\}$ and $\mathscr{Q}'' = \mathscr{Q} \cup \{\exists_{\geq \lceil z_{i^*} \rceil} x\ p_{\text{new}}(x)\}$. These new formulas φ' and φ'' are then dealt with as integer linear problems of larger size. However, if their size exceeds the limit given by Lemma 3.14, the problem is not inserted.

The largest part of the processing in *CQUELBranchAndBound* occurs during the calls to the relaxed solver (line 5), *SolveRelaxedViaColGen*(φ), in which column generation takes place. This process is analogous to that used for PSAT column generation, and it takes as input a CQUEL formula, eventually expanded by the bounding operation and is described in Algorithm 3.3. Its output may contain some non-integral values, but the objective function, which minimizes the solution cost has to be 0 for success to be achieved. Thus *SolveRelaxedViaColGen*(φ) aims at solving the following linear program [4]:

$$\begin{aligned} &\text{minimize } c' \cdot x \\ &\text{subject to } A \cdot x \bowtie b \text{ and } x \geq 0 \end{aligned} \tag{7}$$

Algorithm 3.3 *SolveRelaxedViaColGen*(φ)

Input: A normal form CQUEL formula $\varphi = \langle \mathcal{Q}, \mathcal{U}, \mathcal{E} \rangle$.
Output: A relaxed solution (A, x) , if it exists; or "No", if unsatisfiable.

1: $A_{(0)} = I$; compute cost vector $c^{(0)}$; $x^{(0)} = b$
2: **for** $s = 0$; $c^{(s)\prime} \cdot x^{(s)} > 0$; s++ **do**
3: $\quad z^{(s)} = DualSolution(A_{(s)}, \bowtie b, c^{(s)})$
4: $\quad y^{(s)} = CQUELGenerateColumn(z, \mathcal{U}, \mathcal{E})$
5: $\quad$ **return** "No" **if** column generation failed
6: $\quad A_{(s+1)} = append - column(A_{(s)}, y^{(s)})$
7: $\quad c^{(s+1)} = append(c^{(s)}, 0)$
8: **end for**
9: **return** $A_{(s)}, x^{(s)}$ such that $A_{(s)} x^{(s)} \bowtie b$ // Successful termination

In the linear program (7), {0, 1}-matrix A's columns consist of all possible valuations over $k = |\mathcal{Q}|$ predicates and it has 2^k columns. The cost vector c and solution vector x also have size 2^k, so neither is represented explicitly. Instead, Algorithm 3.3 starts with a square matrix and iterates by generating the columns of A in such a way as to decrease the objective function (lines 2–8).

As Algorithm 3.3 is very similar to the column generation process for PSAT presented by Algorithm 2.1, we only discuss here the main differences between the two.

As we do not have a restriction to "add to one" of PSAT, the initial size of A is $k \times k$, and similarly the cost function c starts with size k and the bound vector b has size $k = |\mathcal{Q}|$. As for the initialization (line 1), A receives the identity matrix I, and the solution x receives b. The initialization of the {0, 1}-cost vector, like in PSAT, is such that $c_j = 1$ iff column A^j is $(\mathcal{U} \cup \mathcal{E})$-unsatisfiable. The added columns will always be $(\mathcal{U} \cup \mathcal{E})$-satisfiable and receive cost 0 (line 7).

As for the similarities, the steps within the loop are exactly the same for both algorithms, and for the same reason. The goal of those steps is to decrease the cost function until it becomes 0, or fail if this is not possible.

The only important difference in the loop is the column generation method. Like in the PSAT case, it uses the dual solution z to compute an inequality based on the reduced cost:

$$z' \cdot y \geq 0 \tag{8}$$

Then it encodes the inequality (8) to a propositional formula, which can be seen as a universal formula over unary predicates $\mathcal{U}'$. It then calls Algorithm 3.1 in the form *JointSAT*$(\mathcal{U} \cup \mathcal{U}', \mathcal{E})$ and if it is satisfiable, returns a valuation for its unary predicates.

Theorem 3.15 *Algorithms 3.2, 3.3 and GenerateColumn provide a decision procedure for the CQUEL-SAT problem.*

Proof (Sketch) The proof is a simplification of the correctness of the branch-and-bound method for ILP [42], due to the fact that CQUEL-SAT requires only a single feasible integer solution instead of searching for optimality in the lattice of feasible integer solutions. Details omitted. □

There is an open source implementation[7] for CQU, that is CQUEL with $\mathscr{E} = \emptyset$. It was developed in C++ and employs an open source linear programming solver[8] and the MiniSAT solver[9] as part of the column generation process. More details can be found at Finger and Bona [21].

3.5 Future Challenges for Counting Quantifiers

We have shown that similar methods can be applied both for Probabilistic Logic and for Counting Quantifiers over unary predicates. Three immediate challenges are suggested by this work.

The first one, which is a direct application of the expansion from CQU to CQUEL, is the application of the counting quantifier techniques developed here to the domain of Description Logics. In particular, it would be nice to have an implementation for CQUEL and its deployment together with the existing tools for Description Logic Reasoning.

The second challenge is more foundational and comes directly from a comparison between results for Probabilistic Logic and Counting Quantifiers, namely, the search for a de Finetti-like notion of coherence for counting quantifiers. In other words, this research topic searches for a betting foundation on counting quantifier statements in analogy to the Probabilistic Logic results described in Sect. 2.2.

The third challenge also comes by analogy with Probabilistic Logic, and it has to do with the existence of inconsistency measures for logic bases involving counting quantifiers. This future investigation may take into consideration that it is possible that the analogy between probabilities and discrete counting breaks at this level, for the simple reason that inconsistency measures for probabilistic bases are continuous and may be approached by convex optimization methods [6], while counting quantifier treatment is discrete and based on integer linear programming techniques, are non-convex.

4 Łukasiewicz Infinitely-Valued Logic and Probabilities

Łukasiewicz Infinitely-valued logic is arguably one of the best studied many-valued logics [12]. It has several interesting properties; semantically, formulas can be seen as taking values in the interval [0, 1]; the semantics is truth functional, so then truth value of compound formula is function of the truth values of its components, and

[7] Available at http://cqu.sourceforge.net.

[8] http://www.coin-or.org/.

[9] http://minisat.se/.

that function is continuous over the interval [0, 1]; in fact, it is piecewise linear. When truth values of propositional symbols are restricted to {0, 1}, the semantics of formulas is that of classical logic; furthermore, it possesses a well developed proof-theory and an algebraic semantics base on MV-algebras.

We present the essentials of Łukasiewicz (always propositional) logic ($Ł_\infty$) and its underlying probability theory. We then introduce the notion of LIP-coherence, which is inspired on de Finetti's notion of a coherent betting system. We define and solve the LIP-satisfiability problem mimicking our analysis of the PSAT and CQUEL-SAT problems.

4.1 Łukasiewicz Infinitely-Valued Logic

Consider a finite set of propositional symbols $\mathscr{P} = \{p_1, \ldots, p_n\}$. We employ $\odot$ and $\oplus$ for Łukasiewicz conjunction and disjunction and write $\neg$ for negation. Usually, only $\neg$ and $\oplus$ are considered basic connectives. So all propositional symbols are formulas and if α and β are formulas in $Ł_\infty$, so are $\neg\alpha$ and $\alpha \oplus \beta$. Define $\alpha \odot \beta$ as $\neg(\neg\alpha \oplus \neg\beta)$ and Łukasiewicz implication $\alpha \to \beta$ as $\neg\alpha \oplus \beta$; it is also,possible to express the lattice connectives $\alpha \wedge \beta$ as $\neg(\alpha \oplus \neg\beta) \oplus \alpha$ and $\alpha \vee \beta$ as $\neg(\neg\alpha \wedge \neg\beta)$.

The semantics of $Ł_\infty$-formulas is given in terms of the rational (or real) interval [0, 1]. A *valuation* is a map $v : \mathscr{P} \to [0, 1]$ which is truth functionally extended to all $Ł_\infty$-formulas in the following way:

$$\begin{aligned} v(\neg\alpha) &= 1 - v(\alpha) \\ v(\alpha \oplus \beta) &= \min(1, v(\alpha) + v(\beta)) \\ v(\alpha \odot \beta) &= \max(0, v(\alpha) + v(\beta) - 1) \end{aligned}$$

The third line above can, of course, be obtained from the definition of $\odot$ in terms of $\neg$ and $\oplus$. Similar truth functional expressions can be obtained for the other connectives:

$$\begin{aligned} v(\alpha \to \beta) &= min(1, 1 - v(\alpha) + v(\beta)) \\ v(\alpha \wedge \beta) &= \min(v(\alpha), v(\beta)) \\ v(\alpha \vee \beta) &= \max(v(\alpha), v(\beta)) \end{aligned}$$

A formula α is *valid* if $v(\alpha) = 1$ for every valuation v, a formula α is *satisfiable* (sometimes called 1-satsfiable) if there exists a v such that $v(\alpha) = 1$; otherwise it is *unsatisfiable*. A set of formulas Γ is satisfiable if there exists a v such that $v(\gamma) = 1$ for all $\gamma \in \Gamma$. If $v(\alpha) = 1$, we say that α is satisfied by v.

It mis easy to see that $\alpha \to \beta$ is satisfied by v iff $v(a) \leq v(b)$. If we define $\alpha \leftrightarrow \beta$ as an abbreviation for $(\alpha \to \beta) \wedge (\beta \to \alpha)$, it follows that $\alpha \leftrightarrow \beta$ is satisfied by v iff $v(\alpha) = v(\beta)$.

4.2 Łukasiewicz Infinitely-Valued Probabilistic Logic and Ł$_\infty$-Coherence

Ł$_\infty$-valuations over propositional symbols $\{p_1, \ldots, p_n\}$ can be seen as points in and n-cube $[0, 1]^n$. To apply the ideas and methods of Quantitative Logic Reasoning to probabilistic Ł$_\infty$, we follow the approach and terminology of Mundici [37]. Define a *convex combination* of a finite set of valuations $v_1, \ldots, v_m$ as a function on formulas into [0, 1] such that

$$C(\alpha) = \lambda_1 v_1(\alpha) + \cdots + \lambda_m v_m(\alpha)$$

where $\lambda_i \geq 0$ and $\sum_{i=1}^{m} \lambda_i = 1$.

In this sense, we define a Łukasiewicz Infinitely-valued Probabilistic (LIP) assignment as an expression of the form

$$\Sigma = \left\{ C(\alpha_i) = q_i \mid q_i \in [0, 1], 1 \leq i \leq k \right\}.$$

The LIP assignment is *satisfiable* if there exists a convex combination C on a set of valuations in the n-cube that jointly verifies all inequalities in Σ. This can be seen in linear algebraic terms as follows. Given a LIP assignment Σ, let $q = (q_1, \ldots, q_k)'$ be the vector of values assigned in Σ, and suppose we are given Ł$_\infty$-valuations $v_1, \ldots, v_m$ and let $\lambda = (\lambda_1, \ldots, \lambda_m)'$ be a vector of C-coefficients. Then consider the $k \times m$ matrix $A = [a_{ij}]$ where $a_i j = v_j(\alpha_i)$. Then Σ is satisfiable if there are $v_1, \ldots, v_m$ and λ such that the set of algebraic constrains (9):

$$\begin{aligned} A \cdot \lambda &= q \\ \sum \lambda_j &= 1 \\ \lambda &\geq 0 \end{aligned} \tag{9}$$

Conditions (9) are analogous to the PSAT constraints in (1).

Note that the number m of columns in A is initially unknown, but the following consequence of Carathéodory's Theorem [16] yields that if (9) has a solution, than it has a "small" solution.

Proposition 4.1 *If a set of restrictions of the form* (9) *has a solution, then there are* $k + 1$ *columns of* A *such that the system* $A_{(k+1)\times(k+1)}\lambda = q_{(k+1)\times 1}$ *has a solution* $\lambda \geq 0$. □

Given a set of pairs of formulas and bets $\langle \alpha_1, q_1 \rangle, \ldots, \langle \alpha_k, q_k \rangle$, we say that there is a Ł$_\infty$-*Dutch book* against the bookmaker (Alice) if the gambler (Bob) can place stakes $\sigma_1, \ldots, \sigma_k \in \mathbb{Q}$ in such a way that, for all valuations v

$$\sum_{i=1}^{k} \sigma_i (q_i - v(\alpha_i)) < 0.$$

Intuitively, in a Dutch Book, Alice's bets $C(\alpha_1), \ldots, C(\alpha_k)$ result in financial disaster for her, for any possible world v.

Definition 4.2 Given a probability assignment to propositional formulas $\{C(\alpha_i) = q_i | 1 \leq i \leq k\}$, the LIP assignment is $Ł_\infty$-*coherent* if there are no Dutch Books against it.

The following extension of de Finetti's Dutch book theorem characterizes coherent LIP-assignments:

Proposition 4.3 (Mundici [36]) *Given a LIP assignment $\Sigma = \{C(\alpha_i) = q_i | 1 \leq i \leq k\}$, the following are equivalent:*

(a) Σ is a $Ł_\infty$-coherent assignment.
(b) Σ is a satisfiable LIP assignment.

It has been shown [9] that the decision problem $Ł_\infty$-coherent LIP-assignments is NP-complete. So, in the case of Łukasiewicz Infinitely-valued ProbabilisticLogic, to decide if a LIP assignment is $Ł_\infty$-coherent, we can again employ linear algebraic methods to solve it. In fact, NP-completeness of LIP satisfiability can be seen as a direct corollary of Proposition 4.1. As Proposition 4.3 asserts that deciding $Ł_\infty$-coherence is the same as determining LIP assignment satisfiability, we refer to this problem as LIPSAT.

4.3 Applying Quantitative Logic Reasoning Methods to LIPSAT

Based on the Quantitative Logic Reasoning approach employed in Sects. 2 and 3, a possible strategy to solve the LIPSAT problem is as follows.

1. Generate a normal form for LIPSAT instances.
2. Provide an algebraic formulation for a normal form LIPSAT.
3. Develop a column generation algorithm based on the algebraic formulation.
4. Implement the algorithm and investigate important empirical properties.

Here we present a development of the first two items. The last two items are currently under progress.

4.4 Algebraic Methods for LIPSAT

In total analogy to PSAT, define a LIPSAT instance as in *(propositional) normal form* if it can be partitioned in two sets, $\langle \Gamma, \Psi \rangle$, where $\Gamma = \{C(\gamma_i) = 1 | 1 \leq i \leq r\}$ and $\Psi = \{C(a_i) = q_i | a_i$ is a propositional symbol, $1 \leq i \leq k\}$, with $0 < q_i < 1$.

The partition Γ is the satisfiable part of the normal form, usually represented only as a set of propositional formulas and Ψ is the *propositional LIP assignment* part. Given a LIP-assignment Σ, it is immediate that there exists a normal form LIPSAT instance $\langle \Gamma, \Psi \rangle$ that is LIP-satisfiable iff Σ is.

Example 4.4 Reconsider Example 2.1 about a doctor who formulates a hypothesis on the need of at least two out of three genes g_1, g_2, g_3 to be active for the disease D to occur. In the classical probabilistic case, it was shown that this hypothesis was inconsistent with the fact that each gene was present in 60% of D-patients.

However, if we model this problem in Łukasiewicz Infinitely-valued Probabilistic-logic, which allows for "partial truths", the hypothesis no longer contradicts the data. In fact, we can have a formulation of the problem directly in normal form, with $\Gamma = \{x_1 \oplus x_2, x_1 \oplus x_3, x_2 \oplus x_3\}$ and $\Psi = \{C(x_1) = C(x_2) = C(x_3) = 0.6\}$.

This LIP assignment has many satisfying pairs of valuations and convex combination. The simplest one contains just one valuation v_1 such that $v_1(x_1) = v_1(x_2) = v_1(x_3) = 0.6$ and $\lambda_1 = 1$. It is immediate that v_1 satisfies all three formulas in Γ and $\lambda_1 v_1$ verifies all three equalities in Ψ. □

The algebraic formalization of LIPSAT (9) when the input LIP assignment is in normal form yields the interesting property that the columns of matrix A can be extended to Γ-satisfying valuations, that is, there is a valuation v over all propositional symbols in Γ such that v satisfies all formulas in Γ and when v is restricted to the symbols $a_1, \ldots, a_k$ in Ψ, it agrees with the respective values in A's column.

This property is used to propose a linear program that allows us to decide the LIP satisfiability of a given LIP assignment. The linear program solves (9) without explicitly representing the large matrix A, using once again a *column generation* method. For that, consider the following linear program:

$$\begin{array}{ll} \min & c' \cdot \lambda \\ \text{subject to} & A \cdot \lambda = q \\ & A\text{'s columns are } a_1, \ldots, a_k \; Ł_\infty - \text{valuations} \\ & \lambda \geq 0 \text{ and } \sum \lambda_i = 1 \end{array} \tag{10}$$

As in Sect. 2.3, the *cost vector* c in (10) is a $\{0, 1\}$-vector such that $c_i = 1$ iff column A^j is Γ-unsatisfying. Thus, the column generation process proceeds by generating Γ-consistent columns. The result of this minimization process reaches total cost $c' \cdot \pi = 0$ iff the input instance is satisfiable, as stated by the following result.

Theorem 4.5 *A normal form LIPSAT instance Σ is LIP-satisfiable iff the corresponding linear program of the form* (10) *terminates with minimal total cost* $c' \cdot \lambda = 0$.

Proof ($\Leftarrow$) If the program terminates, then clearly λ is a convex combination of the columns of A verifying the restriction in Σ.

($\Leftarrow$) If Σ is satisfiable, then by Proposition 4.1 there exists a small k-dimension matrix A and a λ that verifies its restrictions. Note that λ can be seen as a linear combination of the columns of A, which are Ł$_\infty$-valuations by (10); furthermore, $\sum \lambda_i = 1$, so λ is a convex combination of Ł$_\infty$-valuations. As column generation is able to eventually generate cost-decreasing columns, the total cost will reach 0, at which point the program terminates. □

Corollary 4.6 (LIPSAT Complexity) *The problem of deciding the satisfiability of a LIP-assignment is NP-complete.*

Despite the fact that solvable linear programs of the form (10) always have polynomial size solutions, with respect to the size of the corresponding normal form LIP-assignment, the elements of linear program itself (10) may be exponentially large, rendering the explicit representation of matrix A impractical.

Theorem 4.5 serves as a basis for the development of a LIPSAT-solver and its implementation.

4.5 A LIPSAT-Solving Algorithm

The general strategy employed here is similar to that employed to PSAT solving [20, 22], but the column generation algorithm is considerably distinct and requires an extension of Ł$_\infty$-decision procedure.

From the input $\langle \Gamma, \Theta \rangle$, we implicitly deal with matrix A and explicitly obtain the vector of probabilities q mentioned in (10). The basic idea of the simplex algorithm is to move from one feasible solution to another one with a decreasing cost. The pair $\langle B, \lambda \rangle$ consisting of the basis B and a LIP probability distribution λ is a *feasible solution* if $B \cdot \lambda = q$ and $\lambda \geq 0$. We assume that $q_{k+1} = 1$ such that the last line of B forces $\sum_G \lambda_j = 1$, where G is the set of B columns that are Γ-satisfiable. Each step of the algorithm replaces one column of the feasible solution $\langle B^{(s-1)}, \lambda^{(s-1)} \rangle$ at step $s - 1$ obtaining a new one, $\langle B^{(s)}, \lambda^{(s)} \rangle$. The cost vector $c^{(s)}$ is a $\{0, 1\}$-vector such that $c_j^{(s)} = 1$ iff B_j is Γ-unsatisfiable. The column generation and substitution is designed such that the total cost is never increasing, that is $c^{(s)\prime} \cdot \lambda^{(s)} \leq c^{(s-1)\prime} \cdot \lambda^{(s-1)}$.

Algorithm 4.1 presents the top level LIPSAT decision procedure. Lines 1–3 present the initialization of the algorithm. We assume the vector q is in ascending order. Let the D_{k+1} be a $k + 1$ square matrix in which the elements on the diagonal and below are 1 and all the others are 0. At the initial step we make $B^{(0)} = D_{k+1}$, this forces $\lambda_1^{(0)} = q_1 \geq 0$, $\lambda_{j+1}^{(0)} = q_{j+1} - q_j \geq 0$, $1 \leq j \leq k$; and $c^{(0)} = [c_1 \cdots c_{k+1}]'$, where $c_k = 0$ if column j in $B^{(0)}$ is Γ-satisfiable; otherwise $c_j = 1$. Thus the initial state $s = 0$ is a feasible solution.

Algorithm 4.1 main loop covers lines 4–12 which contain the column generation strategy, detailed bellow. If column generation fails the process ends with failure in line 7. Otherwise a column is removed and the generated one is inserted in a process called *merge* at line 9. The loop ends successfully when the objective function (total cost) $c^{(s)\prime} \cdot \lambda^{(s)}$ reaches zero and the algorithm outputs a probability distribution λ and the set of Γ-satisfiable columns in B, at line 13.

Algorithm 4.1 LIPSAT-CG: a LIPSAT solver via Column Generation

Input: A normal form LIPSAT instance $\langle \Gamma, \Theta \rangle$.
Output: No, if $\langle \Gamma, \Theta \rangle$ is unsatisfiable. Or a solution $\langle B, \lambda \rangle$ that minimizes (10).

$q := [\{q_i \mid C(p_i) = q_i \in \Theta, 1 \leq i \leq k\} \cup \{1\}]$ in ascending order;
$B^{(0)} := D_{k+1}$;
$s := 0, \lambda^{(s)} = (B^{(0)})^{-1} \cdot q$ and $c^{(s)} = [c_1 \cdots c_{k+1}]'$;
while $c^{(s)\prime} \cdot \lambda^{(s)} \neq 0$ **do**
  $y^{(s)} = GenerateColumn(B^{(s)}, \Gamma, c^{(s)})$;
  if $y^{(s)}$ column generation failed **then**
    return No; {LIPSAT instance is unsatisfiable}
  else
    $B^{(s+1)} = merge(B^{(s)}, b^{(s)})$
    s++, recompute $\lambda^{(s)}$ and $c^{(s)}$;
  end if
end while
return $\langle B^{(s)}, \lambda(s) \rangle$; {LIPSAT instance is satisfiable}

The procedure *merge* is part of the simplex method which guarantees that given a $k+1$ column y and a feasible solution $\langle B, \lambda \rangle$ there always exists a column j in B such that if $B[j := y]$ is obtained from B by replacing column j with y, then there is λ' such that $\langle B[j := y], \lambda' \rangle$ is a feasible solution.

Column generation method takes as input the current basis B, the current cost c, and the Ł$_\infty$ restrictions Γ; the output is a column y, if it exists, otherwise it signals **No**. The basic idea for column generation is the property of the simplex algorithm called the *reduced cost* of inserting a column y with cost c_y in the basis. The reduced cost r_y is given by

$$r_y = c_y - c' B^{-1} y \tag{11}$$

the objective function is non increasing if $r_y \leq 0$. The generation method always produces a column y thay is Γ-satisfiable so $c_y = 0$. We thus obtain

$$c' B^{-1} y \geq 0 \tag{12}$$

which is an inequality on the elements of y. To force λ to be a convex combination, we make $y_{k+1} = 1$, the remaining elements y_i are valuations of the variables in Θ, so that we are searching for solution to (12) such that $0 \leq y_i \leq 1$, $1 \leq i \leq k$. To finally obtain column y we must extend a Ł$_\infty$-solver that generates valuations satisfying Γ so that it also respects the linear restriction (12). In fact this is not an expressiveness extension of Ł$_\infty$ as the McNaughton property guarantees that (12) is equivalent to some Ł$_\infty$-formula on variables $y_1, \ldots, y_k$ [12]. The practical details on how this can be implemented is detailed in Finger and Preto [23], which also details this final result.

Theorem 4.7 *Consider the output of Algorithm 4.1 with normal form input* $\langle \Gamma, \Theta \rangle$. *If the algorithm succeeds with solution* $\langle B, \lambda \rangle$*, then the input problem is satisfiable* *with distribution λ over the valuations which are columns of B. If the program outputs no, then the input problem is unsatisfiable. Furthermore, there are column selection strategies that guarantee termination.*

5 Conclusion

In this paper we have brought out similarities between three decisions problems in probabilistic logics and counting. The problems deal with satisfiability decision and employ similar linear algebraic methods, fine-tuned for the needs of each specific logic problem. In this way, we belerve that we have elicited grouping them in a class which we named quantitative-logic systems.

There are several other topics were not covered in this work which pertain to all those quantitative logics dealt by this work. Among such topics is the existence of inconsistency measurements, which have been developed for classical probabilistic theories, but not for the other systems. Also, the existence of a phase transition for implementations of the decision procedures described here have been described, but such topic has an empirical nature and thus remains outside the scope of this article.

Acknowledgements We are very thankful to Daniele Mundici for several discussions on multi-valued logics; we would also like to thank two reviewers for their very detailed comments. This work was supported by Fapesp projects 2015/21880-4 and 2014/12236-1 and CNPq grant PQ 306582/2014-7.

References

1. Andersen, K., and D. Pretolani. 2001. Easy cases of probabilistic satisfiability. *Annals of Mathematics and Artificial Intelligence* 33 (1): 69–91.
2. Baader, F., M. Buchheit, and B. Hollander. 1996. Cardinality restrictions on concepts. *Artificial Intelligence* 88 (1): 195–213.
3. Baader, F., S. Brandt, and C. Lutz. 2005. Pushing the EL envelope. In *IJCAI05, 19th International Joint Conference on Artificial Intelligence*, pp. 364–369.
4. Bertsimas, D., and J.N. Tsitsiklis. 1997. *Introduction to Linear Optimization*. Athena Scientific.
5. Biere, A. 2014. Lingeling essentials, a tutorial on design and implementation aspects of the the sat solver lingeling. In *POS@ SAT*, pp. 88. Citeseer.
6. Bona, G.D., and M. Finger. 2015. Measuring inconsistency in probabilistic logic: rationality postulates and dutch book interpretation. *Artificial Intelligence* 227: 140–164.
7. Bona, G.D., F.G. Cozman, and M. Finger. 2014. Towards classifying propositional probabilistic logics. *Journal of Applied Logic* 12(3):349–368. Special Issue on Combining Probability and Logic to Solve Philosophical Problems.
8. Boole, G. 1854. *An Investigation on the Laws of Thought*. London: Macmillan. Available on project Gutemberg at http://www.gutenberg.org/etext/15114.
9. Bova, S., and T. Flaminio. 2010. The coherence of Łukasiewicz assessments is NP-complete. *International Journal of Approximate Reasoning* 51 (3): 294–304.
10. Bulatov, A.A., and A. Hedayaty. 2015. Galois correspondence for counting quantifiers. *Multiple-Valued Logic and Soft Computing* 24 (5–6): 405–424.

11. Calvanese, D., G. De Giacomo, D. Lembo, M. Lenzerini, and R. Rosati (2005). DL-Lite: Tractable description logics for ontologies. In *Proceedings of the 20th National Conference on Artificial Intelligence (AAAI 2005)*, vol. 5, pp. 602–607.
12. Cignoli, R., I. d'Ottaviano, and D. Mundici. 2000. *Algebraic Foundations of Many-Valued Reasoning*, Trends in Logic. Netherlands: Springer.
13. de Finetti, B. 1931. Sul significato soggettivo della probabilità. *Fundamenta Mathematicae* 17(1): 298–329. Translated into English as "On the Subjective Meaning of Probability". In *Probabilitàe Induzione*, ed. P. Monari, and D. Cocchi, 291–321. Bologna: Clueb (1993).
14. de Finetti, B. 1937. La prévision: Ses lois logiques, ses sources subjectives. In *Annales de l'institut Henri Poincaré*, vol. 7:1, pp. 1–68. English translation by Henry E. Kyburg Jr., as "Foresight: Its Logical Laws, its Subjective Sources." In H. E. Kyburg Jr., and H. E. Smokler, "Studies in Subjective Probability", J. Wiley, New York, pp. 93–158, 1964. Second edition published by Krieger, New York, pp. 53–118, 1980.
15. de Finetti, B. 2017. *Theory of probability: A critical introductory treatment*. Translated by Antonio Machí and Adrian Smith. Wiley.
16. Eckhoff, J. 1993. Helly, Radon, and Carathéodory type theorems. In *Handbook of Convex Geometry*, Edited by P.M. Gruber, and J.M. Wills, pp. 389–448. Elsevier Science Publishers.
17. Eén, N. and N. Sörensson. 2003. An extensible SAT-solver. In *SAT 2003*, vol. 2919 *LNCS*, pp. 502–518. Springer.
18. Eén, N., and N. Sörensson. 2006. Translating pseudo-boolean constraints into sat. *Journal on Satisfiability, Boolean Modeling and Computation* 2 (1–4): 1–26.
19. Fagin, R., J.Y. Halpern, and N. Megiddo. 1990. A logic for reasoning about probabilities. *Information and Computation* 87: 78–128.
20. Finger, M., and G.D. Bona. 2011. Probabilistic satisfiability: Logic-based algorithms and phase transition. In *Internatioinal Joint Congerence on Artificial Intelligence (IJCAI)*, Edited by T. Walsh, pp. 528–533. IJCAI/AAAI Press.
21. Finger, M., and G.D. Bona. 2017. Algorithms for deciding counting quantifiers over unary predicates. In *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence, February 4–9, 2017, San Francisco, California, USA.*, Edited by S.P. Singh, and S. Markovitch, pp. 3878–3884. AAAI Press.
22. Finger, M., and G. De Bona. 2015. Probabilistic satisfiability: algorithms with the presence and absence of a phase transition. *Annals of Mathematics and Artificial Intelligence* 75 (3): 351–379.
23. Finger, M., and S. Preto. 2018. Probably half true: Probabilistic satisfiability over Łukasiewicz infinitely-valued logic. In preparation.
24. Georgakopoulos, G., D. Kavvadias, and C.H. Papadimitriou. 1988. Probabilistic satisfiability. *Journal of Complexity* 4 (1): 1–11.
25. Grädel, E., and M. Otto. 1999. On logics with two variables. *Theoretical Computer Science* 224 (1): 73–113.
26. Grädel, E., P.G. Kolaitis, and M.Y. Vardi. 1997. On the decision problem for two-variable first-order logic. *The Bulletin of Symbolic Logic* 3 (1): 53–69.
27. Hailperin, T. 1986. *Boole's Logic and Probability* (Second enlarged edition ed.), vol. 85 *Studies in Logic and the Foundations of Mathematics*. Amsterdam: North-Holland.
28. Hansen, P., and B. Jaumard. 2000. Probabilistic satisfiability. In *Handbook of Defeasible Reasoning and Uncertainty Management Systems*, Edited by J. Kohlas, and S. Moral, vol.5, pp. 321–367. Springer.
29. Hansen, P., B. Jaumard, G.-B.D. Nguetsé, and M.P. de Aragão. 1995. Models and algorithms for probabilistic and bayesian logic. In *IJCAI*, pp. 1862–1868.
30. Horn, A., and A. Tarski. 1948. Measures in boolean algebras. *Transactions of the American Mathematical Society* 64 (3): 467–497.
31. Jaumard, B., P. Hansen, and M.P. de Aragão. 1991. Column generation methods for probabilistic logic. *INFORMS Journal on Computing* 3 (2): 135–148.
32. Kavvadias, D., and C.H. Papadimitriou. 1990. A linear programming approach to reasoning about probabilities. *Annals of Mathematics and Artificial Intelligence* 1: 189–205.

33. Lindström, P. 1966. First order predicate logic with generalized quantifiers. *Theoria* 32 (3): 186–195.
34. Martin, B., F.R. Madelaine, and J. Stacho. 2015. Constraint satisfaction with counting quantifiers. *SIAM Journal on Discrete Mathematics* 29 (2): 1065–1113.
35. Mostowski, A. 1957. On a generalization of quantifiers. *Fundamenta Mathematicae* 44 (2): 12–36.
36. Mundici, D. 2006. Bookmaking over infinite-valued events. *International Journal of Approximate Reasoning* 43 (3): 223–240.
37. Mundici, D. 2011. *Advanced Łukasiewicz calculus and MV-algebras*, Trends in Logic. Netherlands: Springer.
38. Nilsson, N. 1986. Probabilistic logic. *Artificial Intelligence* 28 (1): 71–87.
39. Papadimitriou, C., and K. Steiglitz. 1998. *Combinatorial Optimization: Algorithms and Complexity*. Dover.
40. Pratt-Hartmann, I. 2005. Complexity of the two-variable fragment with counting quantifiers. *Journal of Logic, Language and Information* 14 (3): 369–395.
41. Pratt-Hartmann, I. 2008. On the computational complexity of the numerically definite syllogistic and related logics. *The Bulletin of Symbolic Logic* 14 (1): 1–28.
42. Schrijver, A. 1986. *Theory of Linear and Integer Programming*. New York: Wiley.
43. Walley, P., R. Pelessoni, and P. Vicig. 2004. Direct algorithms for checking consistency and making inferences from conditional probability assessments. *Journal of Statistical Planning and Inference* 126 (1): 119–151.
44. Warners, J.P. 1998. A linear-time transformation of linear inequalities into conjunctive normal form. *Information Processing Letters* 68 (2): 63–69.

Reconciling First-Order Logic to Algebra

Walter Carnielli, Hugo Luiz Mariano and Mariana Matulovic

Abstract We start from the algebraic method of theorem-proving based on the translation of logic formulas into polynomials over finite fields, and adapt the case of first-order formulas by employing certain rings equipped with infinitary operations. This paper defines the notion of M-ring, a kind of polynomial ring that can be naturally associated to each first-order structure and each first-order theory, by means of generators and relations. The notion of M-ring allows us to operate with some kind of infinitary version of Boolean sums and products, in this way expressing algebraically first-order logic with a new gist. We then show how this polynomial representation of first-order sentences can be seen as a legitimate algebraic semantics for first-order logic, an alternative to cylindric and polyadic algebras and closer to the primordial forms of algebraization of logic. We suggest how the method and its generalization could be lifted successfully to n-valued logics and to other non-classical logics helping to reconcile some lost ties between algebra and logic.

1 Introduction

Algebraic logic has emerged as a sub-discipline of Algebra, trying to reflect (in the mirror of algebra) theorems of mathematical logic. As it is well accepted nowadays, George Boole deeply influenced the development of logic by his algebraic approach

W. Carnielli
Department of Philosophy and Centre for Logic, Epistemology and the History of Science, University of Campinas-Unicamp,
Campinas, São Paulo, Brazil
e-mail: walter.carnielli@cle.unicamp.br

H. Luiz Mariano (✉)
Institute of Mathematics and Statistics, University of São Paulo,
São Paulo, Brazil
e-mail: hugomar@ime.usp.br

M. Matulovic
Faculty of Science and Engineering, São Paulo State University (UNESP),
Tupã, São Paulo, Brazil
e-mail: marianamatulovic@tupa.unesp.br

W. Carnielli and J. Malinowski (eds.), *Contradictions, from Consistency to Inconsistency*, Trends in Logic 47, https://doi.org/10.1007/978-3-319-98797-2_13

to logic with the publication of his “The Mathematical Analysis of Logic” of 1847. Boole not only provided a calculus for logic, but offered different interpretations for this calculus. Influenced by Boole’s work, De Morgan introduced the very first calculus of relations.

G. Boole’s attempts were soon to be extended by Charles Peirce and Ernest Schröder. However, a successful leap from propositional calculus to the beginnings of first-order logic had to wait for the developments by Charles Peirce and Gottlob Frege, who independently contribute towards disconnecting the notion of quantifiers from propositional connectives and giving them an appropriate symbolic expression. Peirce used the symbol Σ for the existential quantifier, suggesting a sum, and Π for the universal quantifier, suggesting a product, in certain cases regarding a formula with an existential quantifier as an infinitely long propositional formula.

However, as remarked in [19], the logics considered from 1879 to 1923 (such as those championed by Frege, Peirce, Schröder, Löwenheim, Skolem, Peano, and Russell) were generally richer than the contemporary versions of first-order logic. One of the reasons for such richness was the use of infinitely long expressions especially by Peirce and Schröder (but subsequently also by Hilbert, Löwenheim, and Skolem).

Such infinitely long expressions, understandably, disappeared from contemporary (finitary) first-order predicate calculus. On the other hand, the algebraic approach to propositional classical logic seemed to be naturally extended to first-order logic. The most important steps in this direction are the polyadic algebras of Halmos [15], and the cylindric algebras of Tarski [16, 17], both starting in the 1950s with different emphasis and evoking some alternative approaches such as relation algebras ([17], Chap. 5) and the one presented in [20]. Cylindric algebras are essentially Boolean algebras equipped with unary cylindric operations Cx intended to capture (or to mock) the existential quantifiers $(\exists x)$, while polyadic algebras constitute another approach towards an algebraic representation of first-order logic. It is noteworthy to mention that, concerning modern forms of “algebraizing a Logic”, as expounded in [4], the class of cylindric algebras is essentially the equivalent algebraic semantics of a “tame version” of FOL: See Theorem C.1, p. 71, in [4], for a precise statement. This theorem shows that FOL is algebraizable as a particular case of Blok and Pigozzi’s method, which explores the relationship between metalogical properties enjoyed by a class of logical systems, and their corresponding algebraic properties. As it will be clear, our approach departs radically from this tradition.

However, despite their intrinsic interest and all the papers written on the topic, generalizing these notions and showing connections to other areas, they can hardly be called “natural”: the conceptual difficulties involved and the complex methods of proof make them to stay far away from the earliest algebraic approaches to logic.

Our intention is this paper is, first, to suggest that this disjuncture between first-order logic and its intended algebraic counterpart may be due to the reluctance of modern logicians in using infinitely long expressions of first-order logic, as exemplified by expressing existential quantifiers (essentially infinitary objects) in cylindric algebras by means of the operations Cx. However, infinitary expansions of special mathematical objects (namely, functions) are present in the mathematical milieu at least since their formal introduction by James Gregory and Brook Taylor at the beginning of the seventeenth century. One of the main features of Taylor series is a representation of a function as an infinite sum of terms of lower complexity: the infinite expansion of simpler terms expresses a more complicated function.

Our second intention here is to take charge of this intuition by presenting a treatment of first-order logic by means of formal *infinitary* polynomials, in the intuitive spirit of Taylor series. Such a polynomial representation of logic is shown to be complete, offering a new proof method to first-order logic comparable to the well-known analytic tableaux procedure. We then show how this polynomial representation of first-order formulas could be seen as a legitimate algebraic semantics for first-order logic, alternative to cylindric and polyadic algebras and that is closer to the primeval forms of algebraization of logic.

A recurring point in our research is the distinction between Boolean rings and Boolean algebra. A common confusion is to be deceived by the inter-definability between these structures; in fact, although Boolean algebras and Boolean rings are inter-definable mathematical structures, they are not isomorphic structures (they are just isomorphic categories). It should be noted that the definition of isomorphism requires that isomorphic structures share the same signature (language), which does not occur in this case.

But, more importantly, the inter-definability is maintained just for the bivalent case. For the n-valued case, Boolean rings are generalized in an immediate and natural way via polynomial rings over Galois Field, which does not happen with the Boolean algebras.

The algebraic method of theorem-proving based on the reduction of first-order formulas within certain rings equipped with infinitary operations that we deal with here (where the notion of logical derivability is characterized by the notion of algebraic solubility) constitutes in this way a viable candidate for a new tool for algebraizing logic. The method could be successfully lifted to n-valued logics.

The structure of the paper is the following: Sect. 1.1 presents the polynomial ring calculus as an alternative for algebraization logics, defines the Polynomial Ring Calculus (PRC) for propositional logics and provides some examples of deductions. Section 2 introduces the notion of M-ring,[1] that allows us to operate with some kind of infinitary version of Boolean sums and products. Section 3 shows a polynomial version of First Order Logic (FOL) that completely encodes the notions of Tarskian truth and of derivability inside a first-order theory. Some remarks are offered in Sect. 4, closing this paper.

[1] In this paper "ring" always means commutative ring with unity.

1.1 Polynomial Rings as an Alternative to Algebraizing Logics

What is called “algebraic logic” is a variant of logical reasoning obtained by the association of logic with classes of algebras and other algebraic mathematical structures, relating the properties of logics with algebraic properties of the associated algebras. There is always the risk of confusing this notion with “algebra of logic”, “algebraization of logics” and “algebraic semantics”. The distinctions are very technical and subtle to deal with in an article, but we can at least say that, in its algebraic counterpart, our work can be seen as inserted in the tradition of “algebraic logic”.

The research in algebraic logic proceeds in two different, but often related avenues. One of the approaches studies algebraic structures (or simply algebras) which are relevant to logic(s), e.g. algebras which were obtained from logics. A second approach also studies algebras, e.g. deals with the methodology of solving logical problems by translating them to algebra, solving the algebraic problem and translating the result back to logic.

The algebra of logic (a term coined by Boole) was designed to provide an algorithmic alternative to the traditional approach of Aristotelian logic, as developed in his *Mathematical Analysis of Logic: Being an Essay Towards a Calculus of Deductive Reasoning*, [5].

In Boole’s hands, the addition operation was treated as the union of disjoint sets, obeying the *commutative law*, $x + y = y + x$. The *multiplication* operation was defined as the intersection of two classes, relating with the addition by the *distributive law*, $z(x + y) = zx + zy$, and also obeying the following laws:

- Commutativity: $xy = yx$.
- Index law: $x^2 = x$.

The *Index Law* occupied a central position in the theory of Boole, who considered it a fine example of how a fundamental law of metaphysics, $x^2 = x$, could be seen as just a consequence of the laws of thought. In other words, $x^2 = x \Rightarrow x^2 - x = 0 \Rightarrow x(x - 1) = 0$ represents the law of non-contradiction, that is, the fact that the conjunction of “x” and “not-x” is impossible.

It is interesting to note that, in [5], Boole accepted (in principle) generalization of the index law, e.g, $x^n = x$, for some $n \in \mathbb{N}$. However, in “*The Laws of Thought*” such generalizations were rejected. This happened because of the fact that $x^n = x$ caused him some trouble. As an example, $x^3 = x$ leads to $x^3 - x = 0$, or $x(x^2 - 1) = 0$. In this way, both $x = 1$ and $x = -1$ are roots of this equation. But this posed a problem to Boole’s views, because (-1) is not a valid term in the theory developed by Boole, since it does not obey the index law $(-1)^2 \neq (-1)$.

The fact that prevented the advancement of Boole was not the failure in noticing that the index law could be generalized to $x^n = x$, but the fact that Boole was unable to make sense of it with the mathematical environment of the nineteenth century. In other words, without running the risk of anachronism, we can say that the absence of the concept of algebraic field has prevented Boole from seeing that his algebra of logic could be realized in distinct universes.[2]

In [5], (in Chap. *Of Hypotheticals*, pp. 48–59), Boole defines propositional logic operators in the following polynomial translations:
not X: 1 − x
X and Y: x . y
X or Y (not exclusive): x + y − x . y
X or Y (exclusive): x + y − 2x . y
If X then Y: 1 − x + x . y.

Burris in [7] even says that the algebra of logic developed by Boole was not Boolean algebra. In the nineteenth century the study of symbolic systems for sets was referred to as the "algebra of logic", as Boole called it. In the 1880s C.S. Peirce referred to it as "Boolean algebra", but by that he meant whatever Boole had done. Schröder, a follower of Peirce's work, wrote up an excellent treatment of the subject in 1890 titled *Vorlesungen über die Algebra der Logik*. In this monumental work Schröder refers to Hugh MacColl as one of his most important precursors; see [21] for details. In 1904 Huntington wrote a paper giving several axiom systems for the algebra of logic (for sets). It was in this paper that he considered arbitrary models of the axioms, an important step for the modern subject of Boolean algebra. In 1913, in a follow-up to Huntington's paper, Sheffer [22] referred to the models of any such system of axioms for the algebra of logic as Boolean algebras.

1.1.1 The Method of Polynomial Ring Calculi

The method of Polynomial Ring Calculus (PRC) aims to investigate, at the same time, algebraic notions of proof and new semantic interpretations for classical and non-classical logics. Both, the algebraic versions of proof-theory and the semantic aspects of logic are based on translating logic formulas of a given language $\mathscr{L}$ into polynomial rings, typically a ring $F[X]$ of polynomials in one or more variables X with coefficients in a field F, where the translation is provided by a certain map $(\ )^* : Form(\mathscr{L}) \to F[X]$. The proof method based on polynomial rings is an algebraic proof mechanism that works by performing deductions via polynomial operations on the translated formulas. The elements of the field F represent truth-values, with a given a subset $\emptyset \neq D \subseteq F$ to represent distinguished values (usually $D = \{1\}$). Polynomials may be regarded as the possible truth-values that formulas can take: this makes possible that truth conditions on formulas can be determined by controlling polynomials through certain algebraic operations (the PRC rules). PRC can be regarded as an algebraic semantics, in which the structure of polynomials reflects the structure of truth-value conditions for logic formulas; it can also be seen as a proof method (much as a tableau calculus can be viewed either as a proof-theoretical or as a model-theoretical device).

[2]Distinct, obviously, from Boolean algebras.

Definition 1 (a) A *PRC-interpretation* h of polynomial forms $F[X]$ into a field F is a ring homomorphism which assigns to each polynomial form of the PRC a value in the field F, $h : F[X] \to F$.

(b) A set of polynomials $S \subseteq F[X]$ is *soluble* iff its image under *some* PRC-interpretation is subset of D.

(c) A set of polynomials $S \subseteq F[X]$ is *valid* iff its image under *every* PRC-interpretation is a subset of D. A polynomial $p \in F[X]$ is said to be *invalid* if and only if it is not soluble.

(d) A polynomial $p \in F[X]$ is a *semantic consequence* of the set of polynomials $S \subseteq F[X]$, notation: $S \models\!\approx p$, when for all interpretations $h : F[X] \to F$, if $h[S] \subseteq D$ then $h(p) \in D$.

(e) A polynomial $p \in F[X]$ is a *syntactic consequence* of the set of polynomials $S \subseteq F[X]$, notation: $S \mid\!\sim p$, when from the hypothesis $(s \approx d)$, for each $s \in S$ and some $d \in D$, through the application of *basic rules* to each $s \in S$, can be derived that $(p \approx d)$, for some $d \in D$. Thus a PRC-proof is a (finite) sequence of equations justified by the basic rules of PRC.[3] A polynomial $p \in F[X]$ is a *theorem*, when it is syntactic consequence of the empty set.

(f) A PRC is *compact*, when for each $S \cup \{p\} \subseteq F[X]$, $S \models\!\approx p$ iff exists $S' \subseteq_{fin} S$ such that $S' \models\!\approx p$.

(g) A PRC is *sound*, when for each $S \cup \{p\} \subseteq F[X]$, $S \mid\!\sim p \Rightarrow S \models\!\approx p$.

(h) A PRC is *complete*, when for each $S \cup \{p\} \subseteq F[X]$, $S \models\!\approx p \Rightarrow S \mid\!\sim p$.

(i) A PRC is *semantically adequate*, when for each $\Gamma \cup \{\alpha\} \subseteq Form(\mathscr{L})$, $\Gamma \vDash \alpha \Leftrightarrow \Gamma^* \models\!\approx \alpha^*$.

(j) A PRC is *syntactically adequate*, when for each $\Gamma \cup \{\alpha\} \subseteq Form(\mathscr{L})$, $\Gamma \vdash \alpha \Leftrightarrow \Gamma^* \mid\!\sim \alpha^*$.

We define a particular (p, m)-Polynomial Ring Calculus ((p, m)-PRC) for a given propositional logic system $\mathscr{L}$, based on the Galois Field, $GF(p^m)$ for p prime and m a natural number other than zero, by the following clauses:

1. All the (p, m)-PRC terms are variables, and all its formulas are polynomials in $GF(p^m)[X]$;
2. Operations in (p, m)-PRC are governed by a set of rules. They are:
   a. Index Rules.
      - $p.x \approx 0$, where $(p.x)$ means $(x + x + \cdots + x)$, such addition being performed p times.
      - $x^i.x^j \approx x^k(modq(x))$ in that $q(x)$ is a convenient primitive polynomial that defines $GF(p^m)$, and $k = i + j(mod\ p^m - 1)$.
   b. Ring rules, uniform substitution and Leibniz rules (for equality).[4]

[3]The operator $\mid\!\sim$ clearly induces a *finitary* closure operator $\overline{(\,)} : Parts(F[X]) \to Parts(F[X])$, in particular, $S \mid\!\sim p$ iff exists $S' \subseteq_{fin} S$ such that $S' \mid\!\sim p$.

[4]The symbol $\mid\!\sim$ denotes "reduction by means of polynomial rules"; in order to ease reading, however, we shall use the symbol $\approx$ everywhere when there is no danger of misunderstanding.

In this way, the (p, m)-Polynomial Ring Calculus for a given logic $\mathscr{L}$ (written simply as PRC when there is no danger of confusion) basically consists in translating formulas of $\mathscr{L}$ into polynomials with coefficients in a finite field, and performing deductions through operations (governed by the set of rules defined above) on those polynomials. We say that the polynomial rules *prove* a certain sentence α in $\mathscr{L}$ if its translation in reduced form via application of the rules (the polynomial α^* with coefficients in the Galois field $GF(p^m)$) never outputs values outside the set D of distinguished truth-values.

In summary, defining a concrete PRC for a specific logic $\mathscr{L}$ consists in:

- Selecting a suitable finite field, $GF\ (p^m)$, to represent the truth-values, specifying a subset of distinguished (also called designated in the literature) truth-values.
- Defining a translation function from formulas of $\mathscr{L}$ into polynomials, with variables in the set X and coefficients in $GF\ (p^m)$, namely, $(\ )^* : Form(\mathscr{L}) \rightarrow GF\ (p^m)\ [X]$.
- In certain cases, some constraints on translations will have to be added, as in the cases where modal logic are expressed in polynomial format (see [1]) .

The procedure for obtaining a polynomial representation for a (deterministic or non-deterministic) finite-valued logic begins with the construction of truth-tables for each connective in the language that will be translated. From this point on, in order to characterize the polynomials corresponding to formulas, there are two algorithmic options according to which one may proceed: by means of *Lagrange interpolation* or directly by solving *linear systems* over finite fields.[5]

1.2 The Polynomial Ring Method in Classical Propositional Logic

A PRC for classical propositional logic (or calculus), *CPC* – over the set of propositional variables $\{P_i : i \in \mathbb{N}\}$ – is a *translation* of formulas $FORM\,(CPC)$ in finite polynomials with coefficients in the field $\mathbb{Z}_2$ and variables in the set $X = \{X_i : i \in \mathbb{N}\}$, begins with the *translation* $(\ )^* : Form(CPC) \rightarrow \mathbb{Z}_2[X]$, such that:

- $(P_i)^* = X_i, i \in \mathbb{N}$;
- $(\alpha \wedge \beta)^* = \alpha^*.\beta^*$;
- $(\alpha \vee \beta)^* = \alpha^*.\beta^* + \alpha^* + \beta^*$;
- $(\alpha \rightarrow \beta)^* = \alpha^*.\beta^* + \alpha^* + 1$;
- $(\neg\alpha)^* = \alpha^* + 1$;

Remark 2 Note that the inductive clauses in map $(\ )^*$ are the same clauses that transforms a Boolean ring (BR) over the a set R, $\mathscr{R} = (R, ., +, -, 0, 1)$ into a

[5]It is convenient to show that any finite function can be expressed by means of polynomials over finite fields using two different methods for this: *Lagrange interpolations* and by *solving linear systems*. For more details, see [11, 18].

Boolean algebra (BA) over the same set R, $B(\mathscr{R}) = (R, \wedge, \vee, \to, \neg, 0, 1)$. Remember that, on the other hand, if $\mathscr{B} = (B, \wedge, \vee, \to, \neg, 0, 1)$ is a Boolean algebra, then $R(\mathscr{B}) = (B, ., +, -, 0, 1)$ is a Boolean ring where:

$$a.b := a \wedge b,\ \ a + b := \neg(a \to b) \vee \neg(b \to a),\ \ -a := a.$$

Moreover, a map is a homomorphism of Boolean algebras (respectively, of Boolean rings) iff it is a homomorphism of the associated Boolean rings (respectively, of Boolean algebras). In this way it obtains a pair of inverse isomorphism of categories $BR \leftrightarrows BA$, that "commutes over the category *Set*". □

We will see that the translation of propositional formulas into polynomials is faithful, i.e., the operations on the ring of polynomials corresponding to connectives exactly represent the semantic conditions for classical propositional logic (i.e., faithfully reflects the propositional valuations).

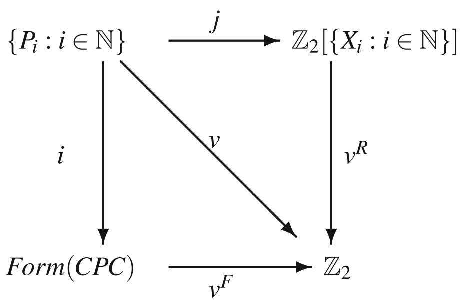
For each valuation $v : \{P_i : i \in \mathbb{N}\} \to \mathbb{Z}_2$, consider:

- $v^F : Form(CPC) \to \mathbb{Z}_2$, the unique extension of v, that is a homomorphism of algebraic structures of type (2,2,2,1) over $C = (\wedge, \vee, \to, \neg)$[6];
- $v^R : \mathbb{Z}_2[\{X_i : i \in \mathbb{N}\}] \to \mathbb{Z}_2$, the unique ring homomorphism such that $v^R(X_i) = v(P_i)$, $i \in \mathbb{N}$.

Then:

Fact 3 It is clear that:

- the mapping $v \mapsto v^F$ establishes a bijection between valuations $\{P_i : i \in \mathbb{N}\} \to \mathbb{Z}_2$ and C-homomorphisms $Form(CPC) \to \mathbb{Z}_2$;
- the mapping $v \mapsto v^R$ establishes a bijection between valuations $\{P_i : i \in \mathbb{N}\} \to \mathbb{Z}_2$ and ring homomorphisms $\mathbb{Z}_2[\{X_i : i \in \mathbb{N}\}] \to \mathbb{Z}_2$.
- for each valuation v, the diagram below commutes, where i and j denote the obvious injective maps.



□

[6]In more details, v^F is recursively defined by: $v^F(P_i) := v(P_i)$, $v^F(\alpha \wedge \beta) := v^F(\alpha) \wedge v^F(\beta)$, $v^F(\alpha \vee \beta) := v^F(\alpha) \vee v^F(\beta)$, $v^F(\alpha \to \beta) := v^F(\alpha) \to v^F(\beta)$, $v^F(\neg\alpha) := \neg v^F(\alpha)$.

Theorem 4 (a) *Let* $v : \{P_i : i \in \mathbb{N}\} \to \mathbb{Z}_2$ *be a valuation, then:* $v^F = v^R \circ (\ )^*$.
(b) *The mapping* $- \circ (\ )^* : Hom_{ring}(\mathbb{Z}_2[\{X_i : i \in \mathbb{N}\}], \mathbb{Z}_2) \to C - Hom(Form(CPC), \mathbb{Z}_2)$, $h \mapsto h \circ (\ )^*$, *is a bijection.*

Proof Item (b) follows from (a) and the bijections in Fact 3. We will now prove $v^F(\psi) = v^R((\psi)^*)$, by induction on the complexity of $\psi \in Form(CPC)$.

- $compl(\psi) = 0$. Then:
  (i) $\psi = P_i$ for some $i \in \mathbb{N}$.
  Then $v^F(P_i) = v(P_i) = v^R(X_i) = v^R(P_i^*)$.
- $compl(\alpha) = n + 1$. Then it holds exactly one of the alternatives below:
  (ii) $\psi = \neg\alpha$, with $compl(\alpha) = n$:
  Then $v^F(\neg\alpha) \underset{def\ v^F}{=} \neg v^F(\alpha) \underset{IH}{=} \neg v^R((\alpha)^*) \underset{def\ \mathbb{Z}_2}{=} 1 + v^R((\alpha)^*) \underset{def\ v^R}{=} v^R(1 + (\alpha)^*) \underset{def(\ )^*}{=} v^R((\neg\alpha)^*)$.
  (iii) $\psi \in \{\alpha \vee \beta, \alpha \wedge \beta, \alpha \to \beta\}$, with $compl(\alpha) + compl(\beta) = n$:
  We will prove only the disjunction case (the other connectives are similar):
  $v^F(\alpha \vee \beta) \underset{def\ v^F}{=} v^F(\alpha) \vee v^F(\beta) \underset{IH}{=} v^R((\alpha)^*) \vee v^R((\beta)^*) \underset{def\ \mathbb{Z}_2}{=} v^R((\alpha)^*) v^R((\beta)^*) + v^R((\alpha)^*) + v^R((\beta)^*) \underset{def\ v^R}{=} v^R((\alpha)^*(\beta)^* + (\alpha)^* + (\beta)^*) \underset{def(\ )^*}{=} v^R((\alpha \vee \beta)^*)$. □

The notions of tautology, valid formulas, etc, on polynomials are defined as in the case of formulas of propositional language, interpreting $1 \in \mathbb{Z}_2$ as *true value* and $0 \in \mathbb{Z}_2$ as the *false* value. Thus:

Corollary 5 *Let* $\Gamma \cup \{\psi\} \subseteq Form(CPC)$.
(a) For each valuation v, $\Gamma \vDash_{v^F} \psi \Leftrightarrow \Gamma^* \approx_{v^R} \psi^*$.
(b) **Semantic adequacy of PRC(CPC)**: $\Gamma \vDash \psi \Leftrightarrow \Gamma^* \approx \psi^*$. □

Fact 6 On the Lindenbaum construction on CPC:

(a) As is well known, the Lindenbaum algebra of CPC,$Lind(CPC) := Form(CPC)/\dashv\vdash$, is a quotient C-algebra that is a Boolean algebra. Then:
$[\alpha] = [\beta] \in Lind(CPC)$ iff $\alpha \dashv\vdash \beta$ iff $\vdash \alpha \leftrightarrow \beta$ iff $\vDash \alpha \leftrightarrow \beta$ iff $\alpha \dashv\vDash \beta$ iff
for all C-homomorphism $V : Form(CPC) \to \mathbb{Z}_2$, $V(\alpha) = V(\beta)$ iff
$(\alpha, \beta) \in \bigcap\{ker(V) :$ for a C-homomorphism $V : Form(CPC) \to \mathbb{Z}_2\}$ iff
for all ring homomorphism $h : \mathbb{Z}_2[X] \to \mathbb{Z}_2$, $h((\alpha)^*) = h((\beta)^*)$.

(b) The inclusion functor $BA \hookrightarrow C - str$ admits a left adjoint functor: i.e., for each C-structure S there is a Boolean algebra $B(S)$ and a C-homomorphism $u_S : S \to B(S)$ with the universal property that, for each a Boolean algebra B and C-homomorphism $h : S \to B$, there is a unique Boolean homomorphism $\widetilde{h} : B(S) \to B$ such that $\widetilde{h} \circ u_S = h$: Consider $E(S) = \bigcap\{E : E$ is a C-congruence relation in S and S/E is a Boolean algebra$\}$, as the subclass of Boolean algebras is closed in $C - str$ under isomorphisms, substructure and products, then $S/E(S)$ is a Boolean algebra and it follows from the "Theorem of Homomorphism" for C-structures that the quotient C-homomorphism $q_S : S \twoheadrightarrow S/E(S)$ has the desired universal property. In particular, as $Form(CPC)$ is the (absolutely) free C-structure over the set $\{P_i : i \in \mathbb{N}\}$, then $Form(CPC)/E(S)$ is the free Boolean algebra over the set $\{P_i : i \in \mathbb{N}\}$ and is also the free Boolean ring over that set.

(c) Let S be an C-structure and consider $E'(S) := \bigcap\{ker(V) :$ for some C-homomorphism $V : S \to \mathbb{Z}_2\}$, then $E'(S) = E(S)$. Since $\mathbb{Z}_2 \in BA$, it is clear that $E(S) \subseteq E'(S)$. As $B(S) = S/E(S)$ is a Boolean algebra, by the Stone representation theorem, the evaluation BA-homomorphism is*injective* $ev_S : B(S) \rightarrowtail \{0, 1\}^{Hom_{BA}(B(S),\{0,1\})}$ which is equivalent to $E(S) = \bigcap\{M : M$ is a C-congruence relation in S and $S/M \cong \mathbb{Z}_2\} = \bigcap\{ker(V) :$ for a C-homomorphism $V : Form(CPC) \to \mathbb{Z}_2\}$. In particular, as $Form(CPC)$ is the (absolutely) free C-structure over the set $\{P_i : i \in \mathbb{N}\}$, then $Lind(CPC) = Form(CPC)/E'(S)$ is the free Boolean algebra over the set $\{P_i : i \in \mathbb{N}\}$ and is also the free Boolean ring over that set.

(d) For each commutative ring with 1, R, denote $I(R)$ the ideal of R generated by the subset $\{a^2 - a : a \in R\}$, then: $R/I(R)$ is a Boolean ring and the quotient homomorphism $q : R \twoheadrightarrow R/I(R)$ has the universal property: for each Boolean ring B and each ring homomorphism $f : R \to B$, there is a unique ring homomorphism $\bar{f} : R/I(R) \to B$ such that $\bar{f} \circ q = f$. By the "Stone representation Theorem for *Boolean rings*", $I(R) = \bigcap\{Z(h) = h^{-1}[\{0\}] :$ for some ring homomorphism $h : R \to \mathbb{Z}_2\}$. If $X = \{X_k : k \in K\}$ then: $I(\mathbb{Z}_2[X]) = \langle\{X_k^2 - X_k : k \in K\}\rangle$, $I(\mathbb{Z}[X]) = \langle\{1 + 1\} \cup \{X_k^2 - X_k : k \in K\}\rangle$ and, as $Lind(CPC)$ is the free Boolean ring over $\{P_k : k \in \mathbb{N}\}$, then $Lind(CPC) \cong \mathbb{Z}_2[X]/I(\mathbb{Z}_2[X]) \cong \mathbb{Z}[X]/I(\mathbb{Z}[X])$.

As a consequence, we have:

Theorem 7 (a) **Compactness of PRC(CPC)**

(b) **Soundeness and Completeness of PRC(CPC)**: *For notion of polynomial proof obtained from the the basic rules*

- *(commutative, with unity) ring rules.*
- *Index rules:*
  $1 + 1 \mathrel{|\!\!\sim} 0$;
  $X_i^2 \mathrel{|\!\!\sim} X_i, i \in \mathbb{N}$.

(c) **Syntactical adequacy of PRC(CPC)**

Proof (a) It is enough to prove: if $S \models\!\approx p$, then there is a finite $S' \subseteq S$ such that $S' \models\!\approx p$. We will prove this contrapositively.

Note that $S \not\models\!\approx p$ iff $S \cup \{1 + p\}$ has a solution iff there exists an interpretation (i.e., a ring homomorphism) $h : \mathbb{Z}_2[X] \to \mathbb{Z}_2$ such that $h[\{1 + s : s \in S\} \cup \{p\}] = \{0\}$ iff there exists an interpretation h such that $\{1 + S\} \cup \{p\} \subseteq Z(h) = h^{-1}[\{0\}]$. As $\mathbb{Z}_2[X]/I(\mathbb{Z}_2[X])$ is a Boolean ring with $I(\mathbb{Z}_2[X]) = \bigcap\{Z(g) :$ for some ring homomorphism $g : \mathbb{Z}_2[X] \to \mathbb{Z}_2\} = \langle\{q^2 - q : q \in \mathbb{Z}_2[X]\}\rangle$, and the maximal ideals in Boolean rings induce a quotient ring $\cong \mathbb{Z}_2$, the maximal ideals M of $\mathbb{Z}_2[X]$ of the form $M = Z(h)$ for some interpretation h are precisely the maximal ideals such that $I(\mathbb{Z}_2[X]) \subseteq M$.

Now suppose that for each $S' \in Parts_{fin}(S)$, it is not the case that $S' \models\!\approx p$, then the subset $\{q^2 - q : q \in \mathbb{Z}_2[X]\} \cup \{1 + S'\} \cup \{p\}$ generates a *proper* ideal of $J(S') \subseteq$

$\mathbb{Z}_2[X]$. As $(Parts_{fin}(S), \subseteq)$ is an upward directed poset, the set $J := \bigcup\{J(S') : S' \in Parts_{fin}(S)\}$ is an ideal of $\mathbb{Z}_2[X]$ that is proper (as $1 \notin J$), thus J can be extended to a maximal ideal M. Note that $\{q^2 - q : q \in \mathbb{Z}_2[X]\} \cup \{1 + S\} \cup \{p\} \subseteq J \subseteq M$, thus $M = Z(h)$ for some ring homomorphism $h : \mathbb{Z}_2[X] \to \mathbb{Z}_2$ and $\{1 + S\} \cup \{p\} \subseteq Z(h)$: this means that it is not the case that $S \approx\!\!\mid p$.

(b) Consider F_X, the absolutely free algebra in the language $R = (., +, -, 0, 1)$ over the set $X = \{X_i : i \in \mathbb{N}\}$, i.e. $|F_X| = \bigcup_{n\in\mathbb{N}} F_n$, where: $F_0 = \{0, 1\} \cup \{X_i : i \in \mathbb{N}\}$ and $F_{n+1} = F_n \cup \{(-, p), (+, p, q), (., p, q) : p, q \in F_n\}$, $|F_X|$ are endowed with the obvious operations. The free (commutative, with 1) ring over the set X is $\mathbb{Z}[X]$ and the ring rules expresses precisely the R-congruence $E(X)$ over F_X such that $\mathbb{Z}[X] \cong F_X/E(X)$. The index rule $(1 + 1 \mid\!\approx 0)$ describes $\mathbb{Z}_2[X]$ as a quotient ring $\mathbb{Z}_2[X] \cong \mathbb{Z}[X]/\langle 1 + 1\rangle$ and the index rule $X_k^2 \mid\!\approx X_k$, $k \in \mathbb{N}$, describes the free Boolean ring $Lind(CPC) \cong \mathbb{Z}_2[X]/I(\mathbb{Z}_2[X])$ over X, where $I(\mathbb{Z}_2[X]) = \langle\{X_k^2 - X_k : k \in \mathbb{N}\}\rangle$.

By the semantic compactness, item (a) above, it is enough to prove $S \mid\!\approx p \Leftrightarrow S \mid\!\sim p$, for each $S \cup \{p\} \subseteq_{fin} \mathbb{Z}_2[X]$. Note that:

- For S finite:

$S \mid\!\approx p$ iff $\{\prod S\} \mid\!\approx p$,[7] because $h[S] = \{1\}$ iff $h(\prod S) = \prod h[S] = 1$, for each ring homomorphism $h : \mathbb{Z}_2[X] \to \mathbb{Z}_2$.

$S \mid\!\sim p$ iff $\{\prod S\} \mid\!\sim p$, because $\{[s] : s \in S\} = \{[1]\} \in \mathbb{Z}_2[X]/I(\mathbb{Z}_2[X])$ iff $[\prod S] = \prod\{[s] : s \in S\} = [1] \in \mathbb{Z}_2[X]/I(\mathbb{Z}_2[X])$.[8]

- For $S = \{s\}$: $s \mid\!\approx p$ iff for all interpretation h, $h(s) = 1$ entails $h(p) = 1$ iff for all interpretation h, $h(sp) = h(s)$ iff for all interpretation h, $h(sp + s) = 0$ iff for all interpretation h, $h(s \to p) = h(sp + s + 1) = 1$ iff $sp + s \in I(\mathbb{Z}_2[X])$ iff $(sp + s) \approx 0$ (iff $(s \to p) \mid\!\sim 1$) iff $([s] \to [p]) = ([s][p] + [s] + [1]) = [sp + p + 1] = [1] \in \mathbb{Z}_2[X]/I(\mathbb{Z}_2[X]) \cong Lind(CPC)$ iff $([s] = [1]$ entails $[p] = 1)$ iff $s \mid\!\sim p$.

(c) For each $\Gamma \cup \{\psi\} \subseteq Form(CPC)$:

$$\Gamma \vdash \psi \underset{sound+complete\ CPC}{\Leftrightarrow} \Gamma \vDash \psi \underset{Corollary\ 5}{\Leftrightarrow} \Gamma^* \mid\!\approx \psi^* \underset{item\ (b)}{\Leftrightarrow} \Gamma^* \mid\!\sim \psi^*. \qquad \square$$

Remark 8 Recall that the quotient map $q : Form(CPC) \twoheadrightarrow Lind(CPC)$ induces a bijective correspondence between the consequence closed subsets $T \subseteq Form(CPC)$ and the ideals $J \subseteq Lind(CPC)$ and that these are in bijective correspondence with the ideals $\bar{J} \subseteq \mathbb{Z}_2[X]$ that contain $I(\mathbb{Z}_2[X])$, we get a "relative version" of translations modulo a CPC-theory $T \subseteq Form(CPC)$, $(\ )^*_T : Form(CPC) \to \mathbb{Z}_2[X]/\bar{J}_T$ into a "relative" ring of polynomials. □

We finish this subsection with some examples:

Example 9 (a) The formula $(P_i \vee P_j \to P_i)$ is satisfiable, i.e. the polynomial $(P_i \vee P_j \to P_i)^*$ has a solution:
Let $h : \mathbb{Z}_2[X] \to \mathbb{Z}_2$ be a ring-homomorphism, then:

$$h((P_i \vee P_j \to P_i)^*) = 1 \Leftrightarrow h((X_i.X_j + X_i + X_j).X_i + (X_i.X_j + X_i + X_j) + 1) = 1 \Leftrightarrow$$

[7] By definition, $\prod S = 1$ whenever $S = \emptyset$.

[8] In a Boolean ring, $ab = 1$ iff $a = b = 1$. Indeed: if $ab = 1$, then $1 = ab = a^2b = a(ab) = a.1 = a$.

$$(h(X_i)h(X_j) + h(X_i) + h(X_j))h(X_i) + (h(X_i)h(X_j) + h(X_i) + h(X_j)) = 0 \Leftrightarrow$$

$$0 = h(X_i)h(X_j) + h(X_j) = (h(X_i) + 1)h(X_j) \Leftrightarrow h(X_i) = 1 \; or \; h(X_j) = 0$$

This means that for the valuations $v : \{P_k : k \in \mathbb{N}\} \to \mathbb{Z}_2$, where $v(P_i) = 1$ or $v(P_j) = 0$, the formula $(P_i \vee P_j \to P_i)$ is satisfiable.

(b) The formula $(P_i \wedge P_j \to P_i)$ is valid.

$$(P_i \wedge P_j \to P_i)^* \;=\; (X_iX_j).X_i + X_iX_j + 1 \;\approx\; X_iX_j + X_iX_j + 1 \;\approx\; 1.$$

This is equivalent to prove that for any assignment of truth values to the variables P_i and P_j, the polynomial $(P_i \wedge P_j \to P_i)^*$ has a solution.

(c) The formula $(P_i \wedge \neg P_i)$ is invalid, because:

Let $h : \mathbb{Z}_2[X] \to \mathbb{Z}_2$ be a ring-homomorphism, then:

$$h((P_i \wedge \neg P_i)^*) = 1 \Leftrightarrow h(X_i(X_i + 1)) = 1 \Leftrightarrow h(X_i).(h(X_i) + 1) = 1 \Leftrightarrow h(X_i)^2 + h(X_i) = 1 \Leftrightarrow 0 = 1$$

This means that there isn't a valuation $v : \{P_k : k \in \mathbb{N}\} \to \mathbb{Z}_2$, that validates the formula $(P_i \wedge \neg P_i)$. □

1.3 The Potential of the PRC

The method of proof by polynomial ring, as expounded in [18], is designed to be a universal method of proof, in the sense of providing a general proof procedure, apt to be used in many different logical systems, such as in propositional many-valued logics (deterministic and non-deterministic), paraconsistent logics, modal logic and First Order Logic.

There are other senses of universality that do not coincide with this, but are related. What is known as "universal logic" is the field of logic that is concerned with investigating what the common characteristics are to all logical structures. Relevant references about this are [2, 3].

The universal logic is not a new logic, but a general theory of logic, considered as mathematical structures. The name was introduced in the 1990s by J.-Y Béziau, but the theme also refers to Alfred Tarski and other Polish logicians such as Adolf Lindenbaum, who have developed a general theory of logic in the late 1920s, based on consequence operators of and matrix logics.

It is also interesting to see the method of polynomial functions as a unifying tool, since it offers a single mathematical object, to wit, polynomials, to compare various aspects of the same logic, as done in [18] with some paraconsistent logics.

Another important aspect of the method is its ability to specify the characteristics of each system to which it is applied. That is, if we are working with truth-functional systems, the polynomials reflect this characteristic; on the other hand, for semitruth-functional logics, new (so-called hidden) variables can be introduced so that the polynomials reflect such characteristics.

Let us consider, as an example, the particular case of three-valued paraconsistent logic of Sette, P_3^1. For this system, we have a set of polynomials in the field resulting $\mathbb{Z}_3$ or ring $\mathbb{Z}_3[X]$, referring to the truth-functional logic three-valued P_3^1, given by:

$$(\neg\alpha)^* = 2x^2 + x + 2;$$
$$(\alpha \to \beta)^* = 2x^2y^2 + x^2 + 2$$

And we also have another polynomial obtained in $\mathbb{Z}_2[X \cup X']$, where X' is a set of hidden variables, resulting from the translation of formulas of the system bivalued reduced from P_3^1, by what is called *Suszko Reduction*. The polynomials in question are:

$$(\neg\alpha)^* = \alpha^* \cdot x_\alpha + 1, \text{ where } x_\alpha \text{ is a hidden variable.}$$
$$(\neg\neg\alpha)^* = (\alpha^* \cdot x_\alpha + 1) \cdot x_\alpha + 1.$$

We can observe that these polynomials have completely different natures, reflecting the fact that two groups of distinct polynomials can be used to characterize a same system.

We could think of using polynomial rings as a new algebraic system which would give a more direct algebraic meaning to several logical systems, in a more natural and intuitive way. This "new algebraization" could then rescue the algebraic side of logical systems, something lost by the contemporary algebraic formalization proposed in the literature, especially for many-valued and paraconsistent logics.[9]

Finally, the polynomial method can be seen as a heuristic device, useful sometimes for discovering new logics or new properties of logic systems, as shown in [12], where the author defines "half-logics and quarter-logics", based on the no-truth-functional connectives developed by Jean-Yves Béziau.

2 *M*-Rings

Let M be an arbitrary set. We introduce the notion of M-ring as a ring equipped with two enumerable families of (infinitary, partial) operators, that allows us to operate with a infinitary version of Boolean products and sums. A structure that is a M-ring for some set M will be called a generically an RNG.

Definition 10 A proto M-ring is a structure $\mathscr{R} := (R, +, \cdot, -, 0, 1, (A_i)_{i\in\mathbb{N}}, (E_i)_{i\in\mathbb{N}})$ where:

- $(R, +, \cdot, -, 0, 1)$ is a structure for the language of rings with unity;

[9]However, it must be noted that a heterodox proposal to algebraize paraconsistent logics is proposed in [6].

- for each $i \in \mathbb{N}$, $A_i, E_i : D_i \to R$ are partial operations on R with commom domain $D_i \subseteq R^M$, such that

$$\mathbf{s} \in D_i \;\Rightarrow\; 1 + \mathbf{s} \in D_i \text{ and } \mathbf{s} + 1 \in D_i$$

(where $1 + \mathbf{s} := (1 + s_a)_{a \in M}$) □

Definition 11 A <u>M-ring</u> $\mathscr{R} := (R, +, \cdot, -, 0, 1, (A_i)_{i \in \mathbb{N}}, (E_i)_{i \in \mathbb{N}})$ is a proto M-ring such that:

- $(R, +, \cdot, -, 0, 1)$ is a commutative ring with unity and characteristic 2;
- for each $i \in \mathbb{N}$ and each $\mathbf{s} \in D_i$, $1 + A_i(\mathbf{s}) = E_i(1 + \mathbf{s})$

□

Remark 12 Let $\mathscr{R}$ be a M-ring and $\mathbf{s} \in R^M$:
(a) $1 + \mathbf{s} = \mathbf{s} + 1$;
(b) $\mathbf{s} \in D_i \;\Leftrightarrow\; 1 + \mathbf{s} \in D_i$;
(c) $E_i(\mathbf{s}) + 1 = A_i(\mathbf{s} + 1)$. □

The fundamental example of M-ring is:

Example 13 For each set M, the set $\mathbb{Z}_2 = \{0, 1\}$ can be endowed with a natural structure of M-ring.

- The structure $\langle \mathbb{Z}_2, +, \cdot, 0, 1 \rangle$ is Boolean ring with unity (i.e., $s^2 = s$), in particular, it is a commutative ring with unity and characteristic 2.
- For each $i \in \mathbb{N}$, take $D_i := \mathbb{Z}_2^M$ and consider the (total) operations given by the usual order structure on $\mathbb{Z}_2$:

$$A_i : \mathbb{Z}_2^M \to \mathbb{Z}_2,\; (s_a)_{a \in M} = \mathbf{s} \mapsto \bigwedge_{a \in M} s_a$$

$$E_i : \mathbb{Z}_2^M \to \mathbb{Z}_2,\; (s_a)_{a \in M} = \mathbf{s} \mapsto \bigvee_{a \in M} s_a$$

Then
$$(1 + \bigwedge_{a \in M} s_a) \;=\; \bigvee_{a \in M} (1 + s_a),$$

because:

$$(1 + \bigwedge_{a \in M} s_a) \;=\; \neg(\bigwedge_{a \in M} s_a) \;=\; \bigvee_{a \in M} (\neg s_a) \;=\; \bigvee_{a \in M} (1 + s_a)$$

□

Remark 14 Clearly, for each set M, the example above can be generalized for each Boolean algebra (or Boolean ring) B and each set M; moreover, if B is a complete Boolean algebra, then its associated M-ring has total operations $A_i, E_i : B^M \to B$, $i \in \mathbb{N}$. □

A key point in this work is the construction, for each L-structure $\mathscr{M}$, of an M-ring denoted by $R(\mathscr{M})$ (with $M = |\mathscr{M}|$) that is a kind of "free" construction given by generators and relations *à la* Lindenbaum algebra construction. Thus we will need a notion of M-homomorphism, and to develop some constructions in the fashion of universal algebra.

Definition 15 Let $\mathscr{R}$ and $\mathscr{R}'$ be proto M-rings. A function $h : R \to R'$ is a M-homomorphism from $\mathscr{R}$ into $\mathscr{R}'$ iff it respects all datas involved. More explicitly:

- $h : (R, +, \cdot, -, 0, 1) \to (R', +', \cdot', -', 0', 1')$ is an homomorphism for the language of rings with unity;
  For each $i \in \mathbb{N}$
- $\mathbf{s} \in D_i \Rightarrow h^M(\mathbf{s}) := ((h(s_a))_{a\in\mathscr{M}} \in D'_i$;
- $h(A_i(\mathbf{s})) = A'_i(h^M(\mathbf{s}))$
- $h(E_i(\mathbf{s})) = E'_i(h^M(\mathbf{s}))$ □

It is clear that the class of proto-M-rings and M-homomorphisms – with obvious composition and identities – is a category, that will be denoted by $proto - M - ring$. We shall write $M - ring$ for the full subcategory of $proto - M - ring$, determined by the subclass of all M-rings. If cBA denote the category of complete Boolean algebras and *complete* homomorphisms, then there is a (faithful) functor $\mathscr{B}^{(M)} : cBA \longrightarrow proto - M - rings$.

Definition 16 Let $\mathscr{P}$ be a proto-M-ring. An M-congruence C in $\mathscr{P}$ is an equivalence relation on P ($C \subseteq P \times P$) such that:

- C is a congruence for the underlying structure $(P, +, \cdot, -, 0, 1)$[10];
- For each $i \in \mathbb{N}$ and $\mathbf{s}, \mathbf{t} \in D_i$, such that $(s_a, t_a) \in C$, for every $a \in M$: $(A_i(\mathbf{s}), A_i(\mathbf{t})) \in C$ and $(E_i(\mathbf{s}), E_i(\mathbf{t})) \in C$. □

The following result is straightforward:

Fact 17 Let $\mathscr{P}$, $\mathscr{P}'$ be proto-M-rings, C be a M-congruence on $\mathscr{P}$ and $h : \mathscr{P} \to \mathscr{P}'$ be a M-homomorphism.

1. The intersection of any family of M-congruences on $\mathscr{P}$ is a M-congruence in $\mathscr{P}$. The set $M - cong(\mathscr{P})$ of all M-congruences in $\mathscr{P}$, endowed with the inclusion relation, is a complete lattice.
2. If $C' \subseteq P' \times P'$ is a M-congruence on $\mathscr{P}'$, then $h^\star(C') := \{(s, t) \in P \times P : (h(s), h(t)) \in C'\}$ is a M-congruence in $\mathscr{P}$. In particular, $ker(h) = \{(s, t) \in P \times P : h(s) = h(t) \in P'\}$ is a M-congruence on $\mathscr{P}$.
3. **M-congruence generated by a relation**: For each relation $S \subseteq P \times P$, there is the least M-congruence relation $\langle S \rangle$ on $\mathscr{P}$ that is above S.

[10] I.e., if $(p, p') \in C$ and $(q, q') \in C$, then: $(-p, -p') \in C$, $(p + q, p' + q') \in C$, $(p.q, p'.q') \in C$.

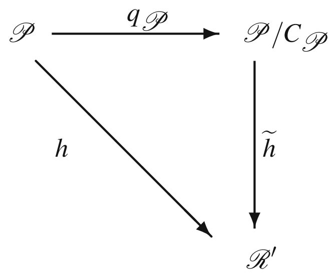
4. **Quotient of proto-M-rings**: There is a unique structure of proto-M-ring on the quotient set P/C, denoted $\mathscr{P}/C$, such that $q_C : \mathscr{P} \twoheadrightarrow \mathscr{P}/C$ is a M-homomorphism and with $\bar{D}_i := (q_C)^M[D_i]$, for each $i \in \mathbb{N}$. Moreover, $\mathscr{P}/C$ is a M-ring whenever $\mathscr{P}$ is a M-ring.
5. **Theorem of M-homomorphism**: The mapping $H \in M - hom(\mathscr{P}/C, \mathscr{P}') \mapsto H \circ q_C \in \{h \in M - hom(\mathscr{P}, \mathscr{P}') : C \subseteq ker(h)\}$ is a bijection. □

As a direct consequence of the Theorem of M-homomorphism, we obtain the following:

Proposition 18 *The (full) subcategory $M - rings \hookrightarrow proto - M - rings$ is reflective. In more detail: for each proto M-ring $\mathscr{P}$, consider $C_{\mathscr{P}}$ the M-congruence generated by the the relation $\sim$*

- *The laws of commutative rings with unity and characteristic 2 are satisfied modulo $\sim$, for instance: $(s + (r + t)) \sim ((s + r) + t)$, $(s.1) \sim s$, $(s.t) \sim (t.s)$, $(s.(r + t)) \sim ((s.r) + (s.t))$, $(s) \sim (-s)$.*
- *For each $i \in \mathbb{N}$ and $\mathbf{s} \in D_i$, $(1 + A_i(\mathbf{s})) \sim (E_i(1 + \mathbf{s}))$.*

Then $\mathscr{P}/C_{\mathscr{P}}$ is a M-ring, and the quotient M-homomorphism $q_{\mathscr{P}} : \mathscr{P} \twoheadrightarrow \mathscr{P}/C_{\mathscr{P}}$ has the universal property:
For each M-ring $\mathscr{R}'$ and each M-homomorphism $h : \mathscr{P} \to \mathscr{R}'$, there is a unique M-homomorphism $\widetilde{h} : \mathscr{P}/C_{\mathscr{P}} \to \mathscr{R}'$, such that $\widetilde{h} \circ q_{\mathscr{P}} = h$, i.e., the following diagram commutes: □



Remark 19 Let $\mathscr{R}$ be a (proto) M-ring, $\mathscr{R}'$ be a (proto) M'-ring and $j : M \longrightarrow M'$ be any function. A function $f : \mathscr{R}' \longrightarrow \mathscr{R}$ (note the reversing) is a *j-morphism* of (proto)RNG when: (i) it preserves $0, 1, -, +, \cdot$; (ii) If $\mathbf{s}' \in D_i' \subseteq \mathscr{R}'^{M'}$, then $f \circ \mathbf{s}' \circ j \in D_i \subseteq \mathscr{R}^M$, $i \in \mathbb{N}$; (iii) If $\mathbf{s}' \in D_i' \subseteq \mathscr{R}'^{M'}$, then $f(A_i'(\mathbf{s}')) = A_i(f \circ \mathbf{s}' \circ j)$ and $f(E_i'(\mathbf{s}')) = E_i(f \circ \mathbf{s}' \circ j)$, $i \in \mathbb{N}$. It is clear that an id_M-morphism is just an ordinary M-ring homomorphism and if $j' : M' \longrightarrow M''$ is a function and $f' : \mathscr{R}'' \longrightarrow \mathscr{R}'$ is a j' morphism, then $f \circ f' : \mathscr{R}'' \longrightarrow \mathscr{R}$ is a $j' \circ j$-morphism. It is easy to check that, in this way, we obtain the category $pRNG$ (respec. RNG), the category of all pRNGs (resp., of all RNGs) with "change of base"-morphisms between them. □

2.1 M-Rings Associated to FO-Languages, FO-Theories and FO-Structures

Let L be a first-order language and M be an arbitrary set; we will define below, by generators and relations: a (free) proto-M-ring $F(M)$ and consider its associated "polynomial" M-ring, where the atomic $L(M)$-formulas are in the role of variables, $R(M) := F(M)/C_{F(M)}$. In particular:
(i) for a certain "canonical" set M^L associated with the language L, we obtain a proto-M^L-ring $\mathscr{F}(L)$ and the M^L-ring $\mathscr{R}(L)$;
(ii) for a certain "canonical" set M^L_T associated with a theory $T \subseteq Form(L)$, we obtain a proto-M^L_T-ring $\mathscr{F}(L)_T$ and the M^L_T-rings $\mathscr{R}(L)_T$;
(iii) to each L-structure $\mathscr{M}$, if $M = |\mathscr{M}|$ is the underlying set of $\mathscr{M}$, we obtain $F(|\mathscr{M}|)$ and $R(|\mathscr{M}|)$.

20 Notations:
(i) By technical convenience, consider first the identifications (= inverse bijections) $Terms(L\dot{\cup}\{a : a \in M\}) \underset{\flat}{\overset{\sharp}{\rightleftarrows}} ClosedTerms(L\dot{\cup}\{a : a \in M\}\dot{\cup}\{k_i : i \in \mathbb{N}\})$, given by $x_i \leftrightsquigarrow k_i$, $i \in \mathbb{N}$.
(ii) We will use the capital letters $\underline{A}$(belard) and $\underline{E}$(loise) to mark the corresponding quantifier symbols $\underline{\forall}$ and $\underline{\exists}$. □

Definition 21 **The set** $|F(M)|$: The underlying set of $F(M)$ is given by a cumulative hierarchy defined by recursion:

- $|F(M)| := \bigcup_{n\in\mathbb{N}} F_n$, where:

1. $F_0 = \{0\} \cup \{1\} \cup \{X_{t_1=t_2}\} \cup \{X_{r(t_1,\dots,t_n)}$: some $n \in \mathbb{N}$ and r a n-ary relational symbol$\}$, where t_i is a closed term in the language $L\dot{\cup}\{a : a \in M\}\dot{\cup}\{k_i : i \in \mathbb{N}\}$.
2. $F_{n+1} = F_n \cup \{\langle -, p\rangle, \langle +, p, q\rangle, \langle \cdot, p, q\rangle, \langle A, \langle S_{i,a}(p)\rangle_{a\in M}\rangle, \langle E, \langle S_{i,a}(p)\rangle_{a\in M}\rangle :$
$p, q \in F_n, i \in \mathbb{N}\}$

- The rank of $p \in |F(M)|$ is the least $j \in \mathbb{N}$ such that $p \in F_j$.
- For each $i \in \mathbb{N}$ and $a \in M$, the compatible family of functions $p \in F_n \mapsto S_{i,a}(p) \in F_n$, with $n \in \mathbb{N}$ – the *substitution of the individual variable* x_i *by* $a \in M$ – is defined, simultaneously with the sequence $(F_j)_{j\in\mathbb{N}}$, by recursion on the rank of p[11]:

1. $S_{i,a}(0) := 0$, $S_{i,a}(1) := 1$.
2. $S_{i,a}(X_{t_1=t_2}) := X_{t_1(k_i|a)=t_2(k_i|a)}$
3. $S_{i,a}(X_{r(t_1,\dots,t_n)}) := X_{r(t_1(k_i|a),\dots,t_n(k_i|a))}$
4. $S_{i,a}(-, p) := (-, S_{i,a}(p))$
5. $S_{i,a}(+, p, q) := (+, S_{i,a}(p), S_{i,a}(q))$
6. $S_{i,a}(\cdot, p, q) := (\cdot, S_{i,a}(p), S_{i,a}(q))$
7. $S_{i,a}(A, \langle S_{j,b}(p)\rangle_{b\in M}) := (A, \langle S_{i,a}(S_{j,b}(p))\rangle_{b\in M})$

[11] Note that the gluing $S_{i,a} : |F(M)| \to |F(M)|$ preserves rank.

8. $S_{i,a}(E, \langle S_{j,b}(p)\rangle_{b\in M}) := (E, \langle S_{i,a}(S_{j,b}(p))\rangle_{b\in M})$ □

Remark 22 By the result above, it becomes clear that the definition of $S_{i,a}(A,\ldots)$ and $S_{i,a}(E,\ldots)$ indeed determine elements of $|F(M)|$. □

Lemma 23 *The family of substitution functions on $|F(M)|$ satisfies:*
(a) $S_{i,a} \circ S_{j,b} = S_{i,b}$, if $i = j$;
(b) $S_{i,a} \circ S_{j,b} = S_{j,b} \circ S_{i,a}$, if $i \neq j$.

Proof By induction on the rank of $p \in |F(M)|$.
We will just provide the proof, when $p = (A, \langle S_{k,c}(q)\rangle_{c\in M})$, for some $q \in |F(M)|$.

- $i = j = k$

$S_{i,a}(S_{j,b}(p)) \underset{def}{=} (A, (S_{i,a}(S_{j,b}(S_{k,c}(q))))_{c\in M}) \underset{IH}{=} (A, (S_{i,a}(S_{j,c}(q))_{c\in M})) \underset{IH}{=} (A, (S_{i,c}(q))_{c\in M}) \underset{IH}{=} (A, (S_{i,b}(S_{k,c}(q))_{c\in M})) \underset{def}{=} S_{i,b}(p)$

- $i = j \neq k$

$S_{i,a}(S_{j,b}(p)) \underset{def}{=} (A, (S_{i,a}(S_{j,b}(S_{k,c}(q))))_{c\in M}) \underset{IH}{=} (A, (S_{i,a}(S_{k,c}(S_{j,b}(q))))_{c\in M}) \underset{IH}{=} (A, (S_{k,c}(S_{i,a}(S_{j,b}(q))))_{c\in M}) \underset{IH}{=} (A, (S_{k,c}(S_{i,b}(q)))_{c\in M}) \underset{IH}{=} (A, (S_{i,b}(S_{k,c}(q)))_{c\in M}) \underset{def}{=} S_{i,b}(p)$

- $i \neq j = k$

$S_{i,a}(S_{j,b}(p)) \underset{def}{=} (A, (S_{i,a}(S_{j,b}(S_{k,c}(q))))_{c\in M}) \underset{IH}{=} (A, (S_{i,a}(S_{j,c}(q)))_{c\in M}) \underset{IH}{=} (A, (S_{j,c}(S_{i,a}(q)))_{c\in M}) \underset{IH}{=} (A, (S_{j,b}(S_{k,c}(S_{i,a}(q))))_{c\in M}) \underset{IH}{=} (A, (S_{j,b}(S_{i,a}(S_{k,c}(q))))_{c\in M}) \underset{def}{=} S_{j,b}(S_{i,a}(p))$

- $j \neq i = k$

$S_{j,b}(S_{i,a}(p)) \underset{def}{=} (A, (S_{j,b}(S_{i,a}(S_{k,c}(q))))_{c\in M}) \underset{IH}{=} (A, (S_{j,b}(S_{i,c}(q)))_{c\in M}) \underset{IH}{=} (A, (S_{i,c}(S_{j,b}(q)))_{c\in M}) \underset{IH}{=} (A, (S_{i,a}(S_{k,c}(S_{j,b}(q))))_{c\in M}) \underset{IH}{=} (A, (S_{i,a}(S_{j,b}(S_{k,c}(q))))_{c\in M}) \underset{def}{=} S_{i,a}(S_{j,b}(p))$

- $i \neq j, j \neq k, k \neq i$

$S_{i,a}(S_{j,b}(p)) \underset{def}{=} (A, (S_{i,a}(S_{j,b}(S_{k,c}(q))))_{c\in M}) \underset{IH}{=} (A, (S_{j,b}(S_{i,a}(S_{k,c}(q))))_{c\in M}) \underset{def}{=} S_{j,b}(S_{i,a}(p))$ □

Definition 24 (The proto-M -ring $F(M)$) The set $|F(M)| := \bigcup_{n\in\mathbb{N}} F_n$ is endowed with an obvious structure of proto-M-ring:
(i) 0, 1 are the constants.
(ii) $- : F(M) \to F(M)$
$p \mapsto \langle -, p\rangle$
(iii) $+ : F(M) \times F(M) \to F(M)$
$(p, q) \mapsto \langle +, p, q\rangle$
(iv) $\cdot : F(M) \times F(M) \to F(M)$
$(p, q) \mapsto \langle \cdot, p, q\rangle$
(v) For each $i \in \mathbb{N}$, let

$$D_i = \left\{ \left\langle S_{i,a}(p) \right\rangle_{a \in M} : p \in F(M) \right\} \subseteq (F(M))^M \ ,$$

then:

$$A_i : D_i \to F(M),$$

$$\left\langle S_{i,a}(p) \right\rangle_{a \in M} \mapsto (A, \left\langle S_{i,a}(p) \right\rangle_{a \in M})$$

$$E_i : D_i \to F(M),$$

$$\left\langle S_{i,a}(p) \right\rangle_{a \in M} \mapsto (E, \left\langle S_{i,a}(p) \right\rangle_{a \in M})$$

By the definition of $S_{i,a}$ above, for each $p \in F(M)$, $1 + S_{i,a}(p) = S_{i,a}(1 + p)$ and $S_{i,a}(p) + 1 = S_{i,a}(p + 1)$, then $\mathbf{s} \in D_i \Rightarrow \{1 + \mathbf{s}, \mathbf{s} + 1\} \subseteq D_i$ and $F(M)$ is indeed a proto-M ring. □

Definition 25 (The M-ring $R(M)$) For each set M, we define the M-ring $R(M)$ as the following quotient set: $R(M) := F(M)/C_{F(M)}$, such that:
(i) $0 := [0]$, $1 := [1]$.
(ii)

$$+ : R(M) \times R(M) \to R(M)$$

$$([p], [q]) \mapsto [p + q]$$

(iii)

$$\cdot : R(M) \times R(M) \to R(M)$$

$$([p], [q]) \mapsto [p \cdot q]$$

(iv)

$$- : R(M) \to R(M)$$

$$[p] \mapsto [-p]$$

(v) For each $i \in \mathbb{N}$, let

$$\bar{D}_i = \left\{ \left\langle [S_{i,a}(p)]_{a \in M} \right\rangle : p \in F(M) \right\} \subseteq (R(M))^M \ ,$$

then:

$$A_i : \bar{D}_i \to R(M),$$

$$\langle [S_{i,a}(p)] \rangle_{a \in M} \mapsto [(A, \langle S_{i,a}(p) \rangle_{a \in M})]$$

$$E_i : \bar{D}_i \to R(M),$$

$$\langle [S_{i,a}(p)] \rangle_{a \in M} \mapsto [(E, \langle S_{i,a}(p) \rangle_{a \in M})]$$

□

Definition 26 Let L be a language and let $T \subseteq Form(L)$ be a first-order theory (i.e., a subset closed under consequence).

(i) Set $M^L := ClosedTerms(L)$. We will denote $\mathscr{F}(L) := F(M^L), \mathscr{R}(L) := R(M^L)$.

(ii) Set $M^L_T := ClosedTerms(L)/\approx_T$, where for each closed terms $u_0, u_1, u_0 \approx_T u_1$ iff $(u_0 = u_1) \in T$. We will denote $\mathscr{F}(L)_T := F(M^L_T), \mathscr{R}(L)_T) := R(M^L_T)$. □

Let $\mathscr{M}$ be an L-structure and consider the set $M := |\mathscr{M}|$. The proto-M-rings $F(M)$ and $R(M)$ can carry more information on the L-structure $\mathscr{M}$ than just on its universe $M = |\mathscr{M}|$: it is natural to consider special M-homomorphism from these proto-M-rings that preserve, in some sense, that additional information. As in the polynomial method for Logics of Formal Inconsistency- LFIs- [13, 18], some restrictions on homomorphisms and constraints on translations are needed (such as [14]). The adequate constraint for obtaining a polynomial ring version of FOL is given by the notion of coherent-homomorphism below.

27 Let $v : \{x_i : i \in \mathbb{N}\} \to M$ be a valuation on $\mathscr{M}$; denote $v_{\mathscr{M}} : Terms(L) \to M$ be the unique extension (defined by recursion on term complexity) of $v : \{x_i : i \in \mathbb{N}\} \to M$; extend $v_{\mathscr{M}}$ to $\widetilde{v} : Terms(L \cup \{a : a \in M\}) \to M$ by $\widetilde{v}(a) = a$, for each $a \in M$; now, by the identifications $Terms(L \dot\cup \{a : a \in M\}) \underset{\flat}{\overset{\sharp}{\rightleftarrows}} ClosedTerms$ $(L \dot\cup \{a : a \in M\} \dot\cup \{k_i : i \in \mathbb{N}\})$, consider $\hat{v} : ClosedTerms(L \dot\cup \{a : a \in M\} \dot\cup \{k_i : i \in \mathbb{N}\}) \to M$ the corresponding function. □

Definition 28 Let $\mathscr{M}$ be a L-structure.

(a) An M-homomorphism, $h : F(\mathscr{M}) \to \mathbb{Z}_2$, is called $\mathscr{M}$-coherent, or simply an $\mathscr{M}$-homomorphism, when:

1. h recovers $\mathscr{M}$, i.e.:
 - for each L-constant symbol c, $h(X_{c=a}) = 1$ iff $c^{\mathscr{M}} = a$;
 - for each n-ary L-functional symbol f, $h(X_{f(a_1,\ldots,a_n)=a}) = 1$ iff $f^{\mathscr{M}}(a_1, \ldots, a_n) = a$;
 - for each n-ary L-relational symbol r, $h(X_{r(a_1,\ldots,a_n)}) = 1$ iff $(a_1, \ldots, a_n) \in r^{\mathscr{M}}$.
2. It induces a valuation v_h on $\mathscr{M}$:
 for each $i \in \mathbb{N}$, there is a unique $a \in M$ such that $h(X_{k_i=a}) = 1$, then define:

$$v_h(x_i) = a \text{ iff } h(X_{k_i=a}) = 1$$

3. The following coherence conditions are satisfied:

$$h(X_{t_1=t_2}) = 1 \Leftrightarrow \widehat{v_h}(t_1) = \widehat{v_h}(t_2) \in M$$

$$h(X_{r(t_1,\ldots,t_n)}) = 1 \Leftrightarrow (\widehat{v_h}(t_1), \ldots, \widehat{v_h}(t_n)) \in r^{\mathscr{M}}$$

(b) An M-homomorphism, $H : R(\mathscr{M}) \to \mathbb{Z}_2$, is called $\underline{\mathscr{M}\text{-coherent}}$, or simply an $\mathscr{M}$-homomorphism, when $H \circ q_{\mathscr{M}} : F(\mathscr{M}) \to \mathbb{Z}_2$ is $\mathscr{M}$-coherent. □

Remark 29 For each L-structure $\mathscr{M}$, we have analogies between: (i) the "free constructions" $Form(L)$ and $F(\mathscr{M})$; (ii) the "realization maps" i.e. valuations $v : \{x_i : i \in \mathbb{N}\} \to M$ and coherent M-homomorphisms $h : F(\mathscr{M}) \to \mathbb{Z}_2$. In the same way that is defined the set $freevar(\varphi) \subseteq \{x_i : i \in \mathbb{N}\}$ of variables that occur free in $\varphi \in Form(L)$, can be defined the set $freekcons(p) \subseteq \{k_i : i \in \mathbb{N}\}$ of k-constants that occur free in $p \in F(\mathscr{M})$. In the same way that is given an inductive proof (on complexity of formulas) that $\mathscr{M} \vDash_{v_1} \varphi \;\Leftrightarrow\; \mathscr{M} \vDash_{v_2} \varphi$, whenever $v_1 \restriction_{freevar(\varphi)} = v_2 \restriction_{freevar(\varphi)}$, can be given an inductive proof (on rank of $p \in F(\mathscr{M})$) that $h_1(p) = h_2(p)$, whenever $h_1(q) = h_2(q)$ for each $q \in F(\mathscr{M})$ with $rank(q) = 0$ such that q occur in p and $k - constants(q) \subseteq freekcons(p)$, in particular, every coherent M-homomorphism into $\mathbb{Z}_2$ coincide on "M-sentences" elements (i.e. on elements $p \in F(\mathscr{M})$ with $freekcons(p) = \emptyset$). □

The result below, corresponds to Fact 3 for FOL, establishes a precise relation between valuations and <u>coherent</u> homomorphisms:

Theorem 30 *Let $\mathscr{M}$ be a L-structure. Then:*

(a) For each valuation v on $\mathscr{M}$, the mapping $h_v : F(\mathscr{M}) \to \mathbb{Z}_2$ defined (by recursion on rank) below, is an $\mathscr{M}$-coherent M-homomorphism:

- $rank(p) = 0$

  (i)

$$h_v(0) := 0 \in \mathbb{Z}_2;$$

$$h_v(1) := 1 \in \mathbb{Z}_2;$$

  (ii)

$$h_v(X_{t_1=t_2}) = 1 \Leftrightarrow \hat{v}(t_1) = \hat{v}(t_2) \in M;$$

$$h_v(X_{r(t_1,\ldots,t_n)}) = 1 \Leftrightarrow \big(\hat{v}(t_1), \ldots, \hat{v}(t_n)\big) \in r^{\mathscr{M}}$$

- $rank(p) = n + 1$

  (iii) - if $rank(r) = n$, $h_v(-r) := -h_v(r)$;

  - if $max\{rank(r), rank(s)\} = n$, $h_v(r + s) := h_v(r) + h_v(s)$, $h_v(r.s) := h_v(r).h_v(s)$;

- *if* $rank(r) = n$, $\mathbf{s} = (S_{i,a}(r))_{a \in M} \in D_i$, $h_v(A_i(\mathbf{s})) := A_i \langle h_v(s_a) \rangle_{a \in M}$, $h_v(E_i(\mathbf{s})) := E_i \langle h_v(s_a) \rangle_{a \in M}$ [12]

(b) The mappings $h \mapsto v_h$ *and* $v \mapsto h_v$ *are inverse bijections between the set of all valuations* $\{x_i : i \in \mathbb{N}\} \to \mathscr{M}$ *and the set of* $\mathscr{M}$*-coherent* M*-homomorphisms* $F(\mathscr{M}) \to \mathbb{Z}_2$.

(c) There is a (induced) bijection between the set of all valuations $\{x_i : i \in \mathbb{N}\} \to \mathscr{M}$ *and the set of* $\mathscr{M}$*-coherent* M*-homomorphisms* $R(\mathscr{M}) \to \mathbb{Z}_2$.

Proof (c) directly follows from (b) and Proposition 18, since $\mathbb{Z}_2$ is an M-ring.

(a) We will prove that $h_v : F(\mathscr{M}) \to \mathbb{Z}_2$ is an $\mathscr{M}$-coherent M-homomorphism by induction on the rank of elements of $F(\mathscr{M})$.

- h_v is an M-homomorphism by the items (i) and (iii) in its recursive definition.
- h_v recovers $\mathscr{M}$, i.e.:
  - for each L-constant symbol c, $h_v(X_{c=a}) = 1 \underset{item\ (ii)}{\Leftrightarrow} \hat{v}(c) = \hat{v}(a) \underset{def.\ \hat{v}}{\Leftrightarrow} c^{\mathscr{M}} = a$;
  - for each n-ary L-functional symbol f, $h(X_{f(a_1,\ldots,a_n)=a}) = 1 \underset{item\ (ii)}{\Leftrightarrow} \hat{v}(f(a_1, \ldots, a_n)) = \hat{v}(a) \underset{def.\ \hat{v}}{\Leftrightarrow} f^{\mathscr{M}}(a_1, \ldots, a_n) = a$;
  - for each n-ary L-relational symbol r, $h(X_{r(a_1,\ldots,a_n)}) = 1 \underset{item\ (ii)}{\Leftrightarrow} \big(\hat{v}(a_1), \ldots, \hat{v}(a_n)\big) \in r^{\mathscr{M}} \underset{def.\ \hat{v}}{\Leftrightarrow} (a_1, \ldots, a_n) \in r^{\mathscr{M}}$.
- It induces a valuation v_{h_v} on $\mathscr{M}$ (moreover $v_{h_v} = v$):
  for each $i \in \mathbb{N}$ and $a \in M$, $h_v(X_{k_i=a}) = 1 \underset{item\ (ii)}{\Leftrightarrow} \hat{v}(k_i) = \hat{v}(a) \underset{def.\ \hat{v}}{\Leftrightarrow} v(x_i) = a$;
  thus there is as unique $a \in M$ such that $h_v(X_{k_i=a}) = 1$ and $v_{h_v}(x_i) = a = v(x_i)$.
- The following coherence conditions are satisfied:
  - $h_v(X_{t_1=t_2}) = 1 \underset{item(ii)}{\Leftrightarrow} \hat{v}(t_1) = \hat{v}(t_2) \underset{item\ above}{\Leftrightarrow} \widehat{v_{h_v}}(t_1) = \widehat{v_{h_v}}(t_2) \in M$;
  - $h(X_{r(t_1,\ldots,t_n)}) = 1 \underset{item(ii)}{\Leftrightarrow} (\hat{v}(t_1), \ldots, \hat{v}(t_n)) \in r^{\mathscr{M}} \underset{item\ above}{\Leftrightarrow} (\widehat{v_{h_v}}(t_1), \ldots, \widehat{v_{h_v}}(t_n)) \in r^{\mathscr{M}}$

(b) In the proof of (a) was established that $v_{h_v} = v$. We will prove now, by induction on the rank of $p \in F(\mathscr{M})$, that $h_{v_h}(p) = h(p)$, for an $\mathscr{M}$-coherent M-homomorphism $h : F(\mathscr{M}) \to \mathbb{Z}_2$.

- $rank(p) = 0$
  (i) As h_{v_h} and h are M-homomorphisms:

$$h_{v_h}(0) = 0 = h(0) \in \mathbb{Z}_2;$$

$$h_{v_h}(1) = 1 = h(1) \in \mathbb{Z}_2;$$

[12] Remember that $rank(S_{i,a}(r) = rank(r)$.

(ii) $h_{v_h}(X_{t_1=t_2}) = 1 \underset{\text{by item (ii) in (a)}}{\Leftrightarrow} \widehat{v_h}(t_1) = \hat{v}_h(t_2) \in M \underset{\text{by coherence conditions on } h}{\Leftrightarrow}$ $h(X_{t_1=t_2}) = 1$;
$h_{v_h}(X_{r(t_1,\ldots,t_n)}) = 1 \underset{\text{by item (ii) in (a)}}{\Leftrightarrow} (\widehat{v_h}(t_1), \ldots, \widehat{v_h}(t_n)) \in r^{\mathscr{M}} \underset{\text{by coherence conditions on } h}{\Leftrightarrow}$ $h(X_{r(t_1,\ldots,t_n)}) = 1$

- $rank(p) = n + 1$
  (iii) As h_{v_h} and h are M-homomorphisms:

$$h_{v_h}(-r) := -h_{v_h}(r) \underset{IH}{=} -h(r) = h(-r);$$

$$h_{v_h}(r + s) := h_{v_h}(r) + h_{v_h}(s) \underset{IH}{=} h(r) + h(s) = h(r + s);$$

$$h_{v_h}(r.s) := h_{v_h}(r).h_{v_h}(s) \underset{IH}{=} h(r).h(s) = h(r.s);$$

$$h_{v_h}(A_i(\mathbf{s})) := A_i \left\langle h_{v_h}(s_a)\right\rangle_{a\in M} \underset{IH}{=} A_i \left\langle h(s_a)\right\rangle_{a\in M} = h(A_i(\mathbf{s}));$$

$$h_{v_h}(E_i(\mathbf{s})) := E_i \left\langle h_{v_h}(s_a)\right\rangle_{a\in M} \underset{IH}{=} E_i \left\langle h(s_a)\right\rangle_{a\in M} = h(E_i(\mathbf{s})).$$

□

We finish this subsection with the notions corresponding to Definition 1 for FOL. As $1 + 1 \approx 0 \in \mathscr{R}(\mathscr{M})$, for each L-structure $\mathscr{M}$, it is natural consider $D = \{1\} \subseteq R(\mathscr{M})$ as the set of designated values.

Definition 31 Let L be a language.

(a) A *PRC(L)-interpretation* of L is an M-homomorphism $\mathscr{M}$-coherent $H : R(\mathscr{M}) \to \mathbb{Z}_2$, for some L-structure $\mathscr{M}$.

(b) A subset $S \subseteq R(\mathscr{M})$ is *soluble* iff its image under *some* PRC(L)-interpretation $H : R(\mathscr{M}) \to \mathbb{Z}_2$ is $D = \{1\}$.

(c) A subset $S \subseteq R(\mathscr{M})$ is *valid* iff its image under *every* PRC(L)-interpretation $H : R(\mathscr{M}) \to \mathbb{Z}_2$ is $\{1\}$. $s \in R(\mathscr{M})$ is said to be *invalid* if and only if it is not soluble.

(d) Let $S \cup \{r\} \subseteq \mathscr{R}(M)$ and $H : R(\mathscr{M}) \to \mathbb{Z}_2$ an interpretation.

Denote $S \mathrel{|\!\!\approx}_{(R(\mathscr{M}),H)} r \Leftrightarrow H[S] = \{1\} \Rightarrow H(r) = 1$. $r \in R(\mathscr{M})$ is *semantic consequence* of the set $S \subseteq R(\mathscr{M})$, notation: $S \mathrel{|\!\!\approx}_{\mathscr{M}} p$, when for all interpretations $H : R(\mathscr{M}) \to \mathbb{Z}_2$, $S \mathrel{|\!\!\approx}_{(R(\mathscr{M}),H)} r$. □

3 A Polynomial Encoding of FOL

Now we define the adequate notion of translation of first-order formulas into M-rings.

Definition 32 Let L be a language and M be a set.

(a) The proto-M-translation is the function defined below by recursion on complexity of L-formulas

$$\tau_M : Form(L) \to F(M)$$

1. $\tau_M(u_1 = u_2) = X_{u_1^\sharp = u_2^\sharp}$[13]
2. $\tau_M(r(u_1, \ldots, u_n)) = X_{(r(u_1^\sharp, \ldots, u_n^\sharp))}$
3. $\tau_M(\varphi \wedge \psi) = \tau_M(\varphi).\tau_M(\psi)$
4. $\tau_M(\varphi \vee \psi) = \tau_M(\varphi).\tau_M(\psi) + \tau_M(\varphi) + \tau_M(\psi)$
5. $\tau_M(\neg\varphi) = \tau_M(\varphi) + 1$
6. $\tau_M(\varphi \wedge \psi) = \tau_M(\varphi).\tau_M(\psi)$
7. $\tau_M(\varphi \to \psi) = \tau_M(\varphi).\tau_M(\psi) + \tau_M(\varphi) + 1$
8. $\tau_M(\forall x_i \varphi) = A_i\left(\left\langle S_{i,a}(\tau_M(\varphi))\right\rangle_{a \in M}\right)$
9. $\tau_M(\exists x_i \varphi) = E_i\left(\left\langle S_{i,a}(\tau_M(\varphi))\right\rangle_{a \in M}\right)$

(b) $\mathscr{T}_M : Form(L) \to R(M)$ is the M-translation iff $\mathscr{T}_M = q_M \circ \tau_M$ for the proto-M-translation $\tau_M : Form(L) \to F(M)$. □

Remark 33 Note that: $varfreevar(\psi) = \emptyset \Rightarrow freekcons(\tau_M(\psi)) = \emptyset$, for each $\psi \in Form(L)$. □

We will need the following technical result:

Lemma 34 *Let $v : \{x_i : i \in \mathbb{N}\} \to M$ be a valuation on $\mathscr{M}$ and denote $v(x_i|a)$ the unique valuation on $\mathscr{M}$ such that $v(x_i|a)(x_i) = a$ and, when $j \neq i$, $v(x_i|a)(x_j) = v(x_j)$. Then for each $p \in F(M)$, $h_{v(x_i|a)}(p) = h_v(S_{i,a}(p))$.*

Proof By induction on the rank of p.

- $rank(p) = 0$

  (i)
  $h_{v(x_i|a)}(0) = 0 = h_v(0) = h_v(S_{i,a}(0))$
  Likewise: $h_{v(x_i|a)}(1) = h_v(S_{i,a}(1))$

  (ii) Note that, by induction of the complexity of $t \in ClosedTerms(L \cup \{a : a \in M\} \cup \{k_i : i \in \mathbb{N}\})$, $\widehat{v(x_i|a)}(t) = \hat{v}(t(k_i|a))$. Then:
  $h_{v(x_i|a)}(X_{t_1=t_2}) = 1 \Leftrightarrow \widehat{v(x_i|a)}(t_1) = \widehat{v(x_i|a)}(t_2) \Leftrightarrow \hat{v}(t_1(k_i|a)) = \hat{v}(t_2(k_i|a)) \Leftrightarrow 1$
  $= h_v(X_{t_1(k_i|a)=t_2(k_i|a)}) = h_v(S_{i,a}(X_{t_1=t_2})$
  Likewise: $h_{v(x_i|a)}(X_{r(t_1,\ldots,t_n)}) = 1 \Leftrightarrow h_v(S_{i,a}(X_{r(t_1,\ldots,t_n)})) = 1$
- $rank(p) = n + 1$

  (iii)
  $h_{v(x_i|a)}(r + s) = h_{v(x_i|a)}(r) + h_{v(x_i|a)}(s) \underset{IH}{=} h_v(S_{i,a(r)}) + h_v(S_{i,a(s)}) = h_v(S_{i,a(r+s)})$
  Likewise: $h_{v(x_i|a)}(-r) = h_v(S_{i,a(1)})(-r)$, $h_{v(x_i|a)}(r.s) = h_v(S_{i,a(r.s)})$.

  (iv) For $p = (A, (\langle S_{j,b}(q)\rangle_{b \in M}))$, for some $q \in F(M)$ with $rank(q) = n$:
  $h_{v(x_i|a)}(A, (\langle S_{j,b}(q)\rangle_{b \in M})) = A_j((h_{v(x_i|a)}(S_{j,b}(q)))_{b \in M}) \underset{IH}{=}$

[13] Remember the identifications in Exercise 27.

$A_j((h_v(S_{i,a}(S_{j,b}(q))))_{b\in M})$
- If $i = j$:
$A_j((h_v(S_{i,a}(S_{j,b}(q))))_{b\in M}) \underset{34.(a)}{=} A_j((h_v(S_{i,b}(q)))_{b\in M})$.
Then $h_v(A, (S_{j,b}(q))_{b\in M}) = 1 \Leftrightarrow A_j((h_v(S_{i,b}(q)))_{b\in M}) = 1 \underset{def\ \mathbb{Z}_2}{\Leftrightarrow}$
$\forall b \in M, h_v(S_{i,b}(q)) = 1 \underset{34.(a)}{=} \forall b \in M, h_v(S_{i,a}(S_{j,b}(q))) = 1 \underset{IH}{\Leftrightarrow}$
$\forall b \in M, h_{v(x_i|a)}(S_{j,b}(q)) = 1 \underset{def\ \mathbb{Z}_2}{\Leftrightarrow} A_j(h_{v(x_i|a)}(S_{j,b}(q))) = 1 \Leftrightarrow$
$h_{v(x_i|a)}(A, (S_{j,b}(q))) = 1$.

- If $i \neq j$:

$h_v(S_{i,a}(A, (S_{j,b}(q))_{b\in M})) = A_j((h_v(S_{i,a}(S_{j,b}(q))))_{b\in M}) = 1 \underset{def\ \mathbb{Z}_2}{\Leftrightarrow}$
$\forall b \in M, h_v(S_{i,a}(S_{j,b}(q))) = 1 \underset{IH}{\Leftrightarrow} \forall b \in M, h_{v(x_i|a)}(S_{j,b}(q)) = 1 \underset{def\ \mathbb{Z}_2}{\Leftrightarrow}$
$A_j(h_{v(x_i|a)}(S_{j,b}(q))_{b\in M}) = 1 \Leftrightarrow h_{v(x_i|a)}(A, (S_{j,b}(q))_{b\in M}) = 1$.
Likewise: $h_{v(x_i|a)}(E, (\langle S_{j,b}(q)\rangle_{b\in M})) = h_v(S_{i,a}(E_j(\langle S_{j,b}(q)\rangle_{b\in M}))$ □

The result above is analogous to Theorem 4 for FOL.

Theorem 35 (Tarski's true in the polynomial form) *Let v be a valuation on $\mathscr{M}$ and φ be an L-formula. Are the following equivalent:*
(a) $\mathscr{M} \vDash_v \varphi$;
(b) $h_v(\tau_M(\varphi)) = 1$;
(c) $\tilde{h}_v([\tau_M(\varphi)]) = 1$

Proof The equivalence between (b) and (c) follows directly from the definitions.
We will establish (a) ⇔ (b) by induction on the complexity of the L-formula φ:

- $compl(\varphi) = 0$

(i)
If $\mathscr{M} \vDash_v (u_1 = u_2) \Leftrightarrow v_{\mathscr{M}}(u_1) = v_{\mathscr{M}}(u_2) \in M \underset{notation\ 27}{\Leftrightarrow} \hat{v}(u_1^\sharp) = \hat{v}(u_2^\sharp) \in M$
$\underset{def\ h_v}{\Leftrightarrow} h_v(X_{u_1^\sharp = u_2^\sharp}) = 1 \underset{def\ \tau_M}{\Leftrightarrow} h_v(\tau_{\mathscr{M}}(u_1 = u_2)) = 1$.
Likewise:
$\mathscr{M} \vDash_v r(u_1, \ldots, u_n) \underset{def\ h_v}{\Leftrightarrow} h_v(X_{r(u_1^\sharp,\ldots,u_n^\sharp)}) = 1 \underset{def\ \tau_M}{\Leftrightarrow} h_v(\tau_{\mathscr{M}}(r(u_1, \ldots, u_n))) = 1$
- $compl(\varphi) = n + 1$

(ii)
If $\varphi = (\neg\alpha)$, with $compl(\alpha) = n$:
$\mathscr{M} \vDash_v (\neg\alpha) \Leftrightarrow \mathscr{M} \nvDash_v \alpha \underset{IH}{\Leftrightarrow} h_v(\tau_{\mathscr{M}}(\alpha)) \neq 1 \Leftrightarrow h_v(\tau_{\mathscr{M}}(\alpha)) = 0 \underset{def\ \mathbb{Z}_2}{\Leftrightarrow} h_v$
$(\tau_{\mathscr{M}}(\alpha)) + 1 = 1 \underset{M-hom}{\Leftrightarrow} h_v(\tau_{\mathscr{M}}(\alpha) + 1) = 1 \underset{def\ \tau_M}{\Leftrightarrow} h_v(\tau_{\mathscr{M}}(\neg\alpha)) = 1$.
If $\varphi = (\alpha \vee \beta)$, with $compl(\alpha) + compl(\beta) = n$:

$\mathscr{M} \vDash_v (\alpha \vee \beta) \Leftrightarrow (\mathscr{M} \vDash_v (\alpha) \text{ or } \mathscr{M} \vDash_v (\beta)) \underset{IH}{\Leftrightarrow} (h_v(\tau_{\mathscr{M}}(\alpha)) = 1 \text{ or } h_v(\tau_{\mathscr{M}}(\beta)) = 1) \underset{def\ \mathbb{Z}_2}{\Leftrightarrow} (h_v(\tau_{\mathscr{M}}(\alpha)).h_v(\tau_{\mathscr{M}}(\beta)) + h_v(\tau_{\mathscr{M}}(\alpha)) + h_v(\tau_{\mathscr{M}}(\beta))) = 1 \underset{M-hom}{\Leftrightarrow} h_v (\tau_{\mathscr{M}}(\alpha).\tau_{\mathscr{M}}(\beta) + \tau_{\mathscr{M}}(\alpha) + \tau_{\mathscr{M}}(\beta)) = 1 \underset{def\ \tau_M}{\Leftrightarrow} h_v(\tau_{\mathscr{M}}(\alpha \vee \beta)) = 1.$

Likewise: if $compl(\alpha) + compl(\beta) = n$, then
$\mathscr{M} \vDash_v \alpha \wedge \beta \Leftrightarrow h_v(\tau_M(\alpha \wedge \beta)) = 1$
$\mathscr{M} \vDash_v \alpha \to \beta \Leftrightarrow h_v(\tau_M(\alpha \to \beta)) = 1.$
(iii)
If $\varphi = \forall x_i \psi$, with $compl(\psi) = n$, then:
$\mathscr{M} \vDash_v \forall x_i \psi \Leftrightarrow \text{for all } a \in M, \mathscr{M} \vDash_{v(x_i|a)} \psi \underset{IH}{\Leftrightarrow} \text{for all } a \in M, h_{v(x_i|a)}(\tau_M(\psi)) = 1 \underset{Lemma\ 34}{\Leftrightarrow} \text{for all } a \in M, h_v(S_{i,a}(\tau_M(\psi))) = 1 \underset{def\ \mathbb{Z}_2}{\Leftrightarrow} A_i((h_v(S_{i,a}(\tau_M(\psi))))_{a \in M}) = 1 \underset{M-hom}{\Leftrightarrow} h_v(A, (S_{i,a}(\tau_M(\psi)))_{a\in M}) = 1 \underset{def\ \tau_M}{\Leftrightarrow} h_v(\tau_M(\forall x_i \psi)) = 1.$

Likewise:
if $\varphi = \exists x_i \psi$, with $compl(\psi) = n$, then $\mathscr{M} \vDash_v \exists x_i \psi \Leftrightarrow h_v(\tau_M(\exists x_i \psi)) = 1.$ □

With the notation in Definition 31, by the combination of Theorems 30.(c) and 35, we obtain our main results:

Theorem 36 (a) *Let $\Delta \subseteq Form(L)$. Then: Δ is satisfiable, if and only if, exists L-structure $\mathscr{M}$ such that $\mathscr{T}_M[\Delta] \subseteq R(M)$ is soluble.*
(b) *For each L-structure $\mathscr{M}$ and each valuation v on $\mathscr{M}$, $\Gamma \vDash_{(\mathscr{M},v)} \psi \Leftrightarrow \mathscr{T}_M[\Gamma] \approx\!\!\!\mid_{(R(M),\tilde{h}_v)} \mathscr{T}_M(\psi)$.*
(c) *For each L-structure $\mathscr{M}$, $\Gamma \vDash_{\mathscr{M}} \psi \Leftrightarrow \mathscr{T}_M[\Gamma] \approx\!\!\!\mid_{R(M)} \mathscr{T}_M(\psi)$.*
(d) **Semantic adequacy of PRC(FOL)**: *$\Gamma \vDash \psi \Leftrightarrow$ (for each $\mathscr{M}$, $\mathscr{T}_M[\Gamma] \approx 1 \Rightarrow$ for each $\mathscr{M}$, $\mathscr{T}_M(\psi) \approx 1$).* □

Remark 37 Let $\mathscr{M}$ be an L-structure, consider a map $b : \{P_i : i \in \mathbb{N}\} \to R(M)$ and denote $B : \mathbb{Z}_2[\{X_i : i \in \mathbb{N}\}] \to R(M)$ the unique ring homomorphism that extend b. Then: $H \in HomCoher(R(M), \mathbb{Z}_2)) \mapsto H \circ B \in Hom(\mathbb{Z}_2[\{X_i : i \in \mathbb{N}\}], \mathbb{Z}_2)$. Thus, for each $S \cup \{p\} \subseteq \mathbb{Z}_2[\{X_i : i \in \mathbb{N}\}]$, $S \approx\!\!\!\mid p \Rightarrow B[S] \approx\!\!\!\mid_{R(M)} B(p)$; in particular for each $r \in R(M)$, $r.r \underset{R(M)}{\approx\!\!\!\mid\!\!\approx} r$: this means that the "index rule" $r.r \approx r$[14] can be added as "correct rule" in PRC(FOL).[15]

For each map $a : \{P_i : i \in \mathbb{N}\} \to Form(L)$, consider $b_a := \mathscr{T}_M \circ a : \{P_i : i \in \mathbb{N}\} \to R(M)$ and denote:

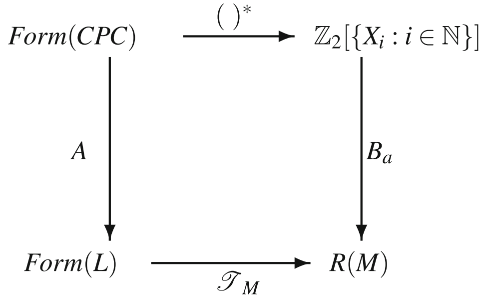
$A : Form(CPC) \to Form(L)$, the recursively defined (unique) extension of a;
$B_a : \mathbb{Z}_2[\{X_i : i \in \mathbb{N}\}] \to R(M)$, the unique ring homomorphism that extends b_a.

---

[14]This rule is equivalent to $r \to r \approx 1$, since $(r \to r) = r.r + r + 1$ and $R(M)$ is, by construction, a ring of characteristic 2, i.e. the "index rule" $r + r = 0$ is already true in $R(M)$.

[15]I.e., for each M-homomorphism $\mathscr{M}$-compatible $H : R(M) \to \mathbb{Z}_2$, the left and the right side of the rule have the same image under H.

It can be proved by induction on the complexity of $\alpha \in Form(CPC)$ that the following diagram commutes:



Thus each logical map of CPC into FOL, $A : Form(CPC) \to Form(L)$, has a polynomial extension $B_a : \mathbb{Z}_2[\{X_i : i \in \mathbb{N}\}] \to R(M)$. Moreover for $\Gamma \cup \{\psi\} \subseteq Form(CPC)$, the implication $\Gamma \vDash^{CPC} \psi \;\; \Rightarrow \;\; A[\Gamma] \vDash^{FOL} A(\psi)$, has a polynomial version $(\Gamma)^* \approx\!\!\!\models \psi^* \;\Rightarrow\; B_a[(\Gamma)^*] \approx\!\!\!\models_{R(M)} B_a(\psi^*)$. □

Let us now provide some examples:

Example 38 1. For each L-structure $\mathscr{M}, p, q, r_1, \ldots, r_n \in F(\mathscr{M}), i, j \in \mathbb{N}$, the following "rules are $R(M)$-correct":

(a) $(A, S_{i,a}(A, (S_{j,b}(p))_{b\in M})_{a\in M}) = (A, S_{j,b}(A, (S_{i,a}(p))_{a\in M})_{b\in M})$, by Definition 21 and Lemma 23. In particular, for each L-formula $\psi(x_i, x_j)$, the L-formula

$$\forall x_i \forall x_j \psi(x_i, x_j) \leftrightarrow \forall x_j \forall x_i \psi(x_i, x_j),$$

is valid. Note that is enough to show that $\tau_M(\textit{left side}) = \tau_M(\textit{right side})$:

$$\tau_M(\forall x_i \forall x_j \psi(x_i, x_j)) \;=_{df.27}\; A_i(S_{i,a}(\tau_M(\forall x_j \psi(x_i, x_j)))_{a\in M}) =_{df.27}$$

$$A_i(S_{i,a}(A_j(S_{j,b}(\tau_M(\psi(x_i, x_j)))_{b\in M}))_{a\in M}) \;=_{df.17}$$

$$A_i(A_j(S_{i,a}(S_{j,b}(\tau_M(\psi(x_i, x_j)))_{b,a\in M}))) =_{lem.19} A_i(A_j(S_{j,b}(S_{i,a}(\tau_M(\psi(x_i, x_j)))_{a,b\in M}))) =_{df.20}$$

$$A_j(A_i(S_{j,b}(S_{i,a}\tau_M(\psi(x_i, x_j))_{a,b\in M}))) \;=\; \tau_M(\forall x_j \forall x_i \psi(x_i, x_j))$$

(b) $[(A, (S_{i,a}(p)))_{a\in M}].[(A, (S_{i,b}(q)))_{b\in M}] \approx [(A, (S_{i,c}(p.q)))_{c\in M}]$: by a direct calculation with binary and infinitary infs;

(c) "Infinite products are closed under finite sums":
$[(A, (S_{i,a}(p)))_{a\in M}].[(A, (S_{i,b}(r_1)))_{b\in M} + \cdots (A, (S_{i,b}(r_n)))_{b\in M}]$
$\approx [(A, (S_{i,c}(p.r_1)))_{c\in M} + \cdots + (A, (S_{i,c}(p.r_n)))_{c\in M}]$: by (b) above and the distributive law in the ring $R(M)$.

2. $\forall x_i \varphi(x_i) \to \exists x_i \varphi(x_i)$ is a valid L-formula, by the PRC(FOL). Indeed, we will show that for each L-structure $\mathscr{M}$, the M-translation of this

L-formula, $\mathcal{T}_M(\forall x_i\varphi(x_i) \to \exists x_i\varphi(x_i)) \in R(M)$, can be reduced to 1 by the algebraic laws satisfied by $R(M)$ and some applications of correct rules:
$\mathcal{T}_M(\forall x_i\varphi(x_i) \to \exists x_i\varphi(x_i)) =_{df.32} \mathcal{T}_M(\forall x_i\varphi(x_i)).\mathcal{T}_M(\exists x_i\psi(x_i))+$
$\mathcal{T}_M(\forall x_i\varphi(x_i)) + 1 =_{df.32}$
$A_i(S_{i,a}(\mathcal{T}_M(\varphi(x_i)))_{a\in M}).\big(E_i(S_{i,a}(\mathcal{T}_M(\varphi(x_i)))_{a\in M})\big) +$
$A_i(S_{i,a}(\mathcal{T}_M(\varphi(x_i)))_{a\in M}) + 1 =_{df.11}$
$A_i(S_{i,a}(\mathcal{T}_M(\varphi(x_i)))_{a\in M}).(1 + A_i(1 + S_{i,a}(\mathcal{T}_M(\varphi(x_i))))_{a\in M})+$
$A_i(S_{i,a}(\mathcal{T}_M(\varphi(x_i)))_{a\in M}) + 1 =$[16]
$A_i(S_{i,a}(\mathcal{T}_M(\varphi(x_i)))_{a\in M}).A_i(1 + S_{i,a}\mathcal{T}_M(\varphi(x_i)))_{a\in M}) + 1 =_{df.21}$
$A_i(S_{i,a}(\mathcal{T}_M(\varphi(x_i))))_{a\in M}.A_i(S_{i,a}(1 + \mathcal{T}_M(\varphi(x_i)))_{a\in M}) + 1 \approx_{correct\ rule\ in\ 1.(b)}$
$A_i(S_{i,a}(\mathcal{T}_M(\varphi(x_i)).(1 + \mathcal{T}_M(\varphi(x_i))))_{a\in M}) + 1 \approx$[17] $0 + 1 = 1.$

3. By the use of the "correct rules" to make shortcuts, the PRC(FOL) can establish the validity of the L-formulas:
   (a) $\exists x_i\forall x_j\psi(x_i, x_j) \to \forall x_j\exists x_i\psi(x_i, x_j)$;
   (b) $\forall x_i(\alpha(x_i) \to \beta(x_i)) \to (\forall x_i\alpha(x_i) \to \forall x_i\beta(x_i))$. □

Remark 39 The above polynomial encoding of Tarskian semantics is compatible with the elementary embeddings of L-structures. More precisely, let $j : \mathcal{M} \longrightarrow \mathcal{M}'$ be an elementary embeddings of L-structures. Then the proto-M'-ring $F(M')$ has an underlying proto-$j[M]$-ring, $F(M')_M$ given (recursively) by functions $U_n : F_n(M') \longrightarrow F_n(M')_M$, $n \in \mathbb{N}$, satisfying obvious conditions and $U_{n+1}(A, (S_{i,a'}(p'))_{a'\in M'}) = (A, (S_{i,j(a)}(U_n(p')))_{j(a)\in j[M]})$. Clearly, $j : M \xrightarrow{\cong} j[M]$ induces a kind of "relative" $(M, j[M])$-isomorphism of proto-rings, $F(M')_M \xrightarrow{\cong} F(M)$; composing this bijection with the map $U : F(M') \longrightarrow F(M')_M$ above, we obtain an induced map $U_j : F(M') \longrightarrow F(M)$ (note the reversing of arrows). This map $U_j : F(M') \longrightarrow F(M)$ factors uniquely through the projections $q : F(M) \twoheadrightarrow R(M)$ and $q' : F(M') \twoheadrightarrow R(M')$ given an well defined map $\bar{U}_j : R(M') \longrightarrow R(M)$. U_j and $\bar{U}_j$ are j-morphisms (see Remark 19). If $L - struc^{\preccurlyeq}$ denotes the category of L-structures and elementary embeddings, we get, in this way, two *contravariant* functors $U : L - struc^{\preccurlyeq} \longrightarrow pRNG$ and $\bar{U} : L - struc^{\preccurlyeq} \longrightarrow RNG$.

Concerning (proto) translations, it is straightforward to check that

$$U_j \circ \tau_{M'} = \tau_M \ and\ \bar{U}_j \circ \mathcal{T}_{M'} = \mathcal{T}_M.$$

If $v : \{x_i : i \in \mathbb{N}\} \longrightarrow M$ is a valuation in $\mathcal{M}$, then the equivalence

$$\mathcal{M} \vDash_v \phi(\bar{x}) \Leftrightarrow \mathcal{M}' \vDash_{j\circ v} \phi(\bar{x})$$

is encoded by the equivalent equations

$$h_v \circ \tau_M = h_{j\circ v}\tau_{M'}\ ;\ \tilde{h}_v \circ \mathcal{T}_M = \tilde{h}_{j\circ v}\mathcal{T}_{M'}.$$

[16]By the distributive law in $R(\mathcal{M})$ and $r + r = 0$.

[17]By the correct rule $r.(1 + r) \approx 0$.

Remark 40 The "syntactic" map $(\ )^* : Form(CPC) \to \mathbb{Z}_2[X]$ gives a (free) ring version of the absolutely free algebra $Form(CPC)$; we saw in Fact 6 and Theorem 7, that the "semantical reduction" of this map establishes an isomorphism of (free) Boolean rings $Lind(CPC) \cong \mathbb{Z}_2[X]/I(\mathbb{Z}_2[X])$ that encodes a sound and complete proof-theoretic counterpart of the logical and polynomial version of CPC.

It is natural to pose the question of whether there is a sound and complete polynomial version of FOL for each language L that is encoded by a semantical reduction of a syntactical map. In particular, is there a "privileged" map $\mathscr{T}_M : Form(L) \to R(M)$? The word "privileged" may suggest a category-theoretic interpretation in the style of "satisfying a universal property".

For each L-structure $\mathscr{M}$, there is an well-defined function $i^{\mathscr{M}} : M^L = ClosedTerms(L) \longrightarrow |\mathscr{M}|$ and, if $\mathscr{M}$ is a model of a theory $T \subseteq Form(L)$, then there is an induced function on the quotient set $i^{\mathscr{M}}_T : M^L_T = ClosedTerms(L)/\approx_T \longrightarrow |\mathscr{M}|$; see Definition 26 and the notation therein. There are (recursively defined) induced functions $I^{\mathscr{M}} : F(M^L) \longrightarrow F(|\mathscr{M}|)$ and $\bar{I}^{\mathscr{M}} : R(M^L) \longrightarrow R(|\mathscr{M}|)$ given by "inclusion of atomic expressions" and, concerning translations, they satisfy the equations:

$$I^{\mathscr{M}} \circ \tau_{M^L} = \tau_{|\mathscr{M}|} \ , \ \bar{I}^{\mathscr{M}} \circ \mathscr{T}_{M^L} = \mathscr{T}_{|\mathscr{M}|}.$$

A "*mod T*" relative statement can be given and, concerning translations, holds for an analogous pair of equations.

It is natural consider the M^L-congruence $J^L := \bigcap\{ker(H \circ I^{\mathscr{M}}) :$ for some M-homomorphism $\mathscr{M}$-coherent $H : R(M) \to \mathbb{Z}_2$ and some L-structure $\mathscr{M}\}$. We can see that J^L has a role analogous to the ideal $I(\mathbb{Z}_2[X]) \subseteq \mathbb{Z}_2[X]$ concerning a definition of a sound and complete polynomial version of the first-order theory over the language L. We can consider also a "*mod T*" relative statement concerning the M^L-congruence $J^L_T := \bigcap\{ker(H \circ I^{\mathscr{M}}) :$ for some M-homomorphism $\mathscr{M}$-coherent $H : R(M) \to \mathbb{Z}_2$ and some L-structure $\mathscr{M}$ such that M satisfies $T\}$. □

4 Final Remarks and Future Work

This paper has offered some new and, we hope, interesting views on the role of algebraizing logics based on the uses of formal polynomials. This section makes some observations on the paper's key points, listing firstly some more technical remarks.

Remark 41 The encoding of Tarski's true in the M-rings $R(\mathscr{M})$ is such that allows us to transfer model-theoretical methods of FOL into this algebraic setting. More precisely, one of the main characteristics of model theory of FOL is the (useful) variation of languages, in particular, the addition of new constants to a language: this aspect can be faithfully represented in the algebraic theory developed above. For instance, let $j : L \to L'$ be a morphism of languages (i.e. $j = (j_C, j_F, j_R)$ where $j_C : C \to C'$, $j_F : F^{(n)} \to F'^{(n)}$, $j_R : R^{(n)} \to R'^{(n)}, n \in \mathbb{N}$). Then denote: $j_T : Terms(L) \to Terms(L')$ the (recursively defined) induced map on terms; $j_F : Form(L) \to Form(L')$ the (recursively defined) induced map on formulas. Consider the map between classes[18] $j^\star : L' - struc \to L - struc$ such that for each L'-structure $\mathscr{M}'$:

- $|j^\star(\mathscr{M}')| = |\mathscr{M}'| := M'$;
- $c \in C \mapsto c^{j^\star(\mathscr{M}')} := (j_C(c))^{\mathscr{M}'} \in M'$;
- $f \in F^{(n)} \mapsto f^{j^\star(\mathscr{M}')} := (j_F(f))^{\mathscr{M}'} : M'^n \to M'$;
- $r \in R^{(n)} \mapsto r^{j^\star(\mathscr{M}')} := (j_R(r))^{\mathscr{M}'} \subseteq M'^n$.

Let $v : \{x_i : i \in \mathbb{N}\} \to M'$ and consider the (recursively defined) extensions: $v_{\mathscr{M}'} : Terms(L') \to M'$, $v_{j^\star(\mathscr{M}')} : Terms(L) \to M'$. Then:

- $v^{\mathscr{M}'}(j_T(u)) = v^{j^\star(\mathscr{M}')}(u)$, for each $u \in Terms(L)$;
- $\mathscr{M}' \vDash_v^{L'} j_F(\psi) \Leftrightarrow j^\star(\mathscr{M}') \vDash_v^L \psi$, for each $\psi \in Form(L)$.

With notations in Exercise 27, by Theorem 35 the (meta-logical) equivalence in the item just above corresponds to the (algebraic) equation:
(EQ) $h_v^{\mathscr{M}'}(\tau_{\mathscr{M}'}(j_F(\psi))) = h_v^{j^\star(\mathscr{M}')}(\tau_{j^\star(\mathscr{M}')}(\psi))$, for each $\psi \in Form(L)$.

Moreover, as $|j^\star(\mathscr{M}')| = |\mathscr{M}'| := M'$, both constructions $F(j^\star(\mathscr{M}'))$, $F(\mathscr{M}')$ are proto-M'-rings. Denote $j_T^+ : ClosedTerms(L \cup M' \cup K) \to ClosedTerms(L' \cup M' \cup K)$ the induced map on (extended, closed) terms. Let $j^F : F(j^\star(\mathscr{M}')) \to F(\mathscr{M}')$ the *unique* M'-homomorphism such that, for each $t_i \in ClosedTerms(L \cup M' \cup K)$:

$$X_{t_1=t_2} \mapsto X_{j_T^+(t_1)=j_T^+(t_2)};$$
$$X_{r(t_1,\dots,t_n)} \mapsto X_{j_R(r)(j_T^+(t_1),\dots,j_T^+(t_n))},\ r \in R^{(n)},\ n \in \mathbb{N}.$$

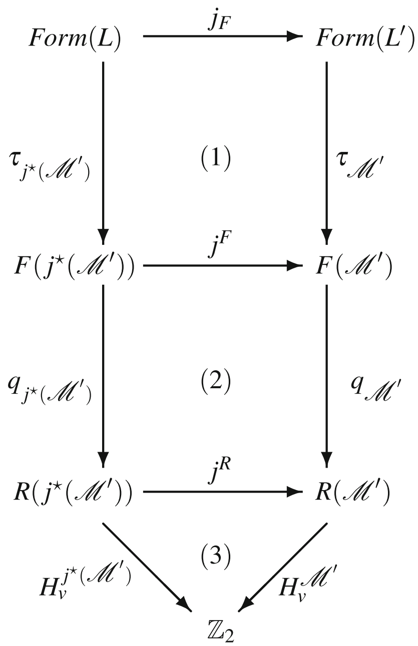
By induction on complexity of L-formulas, can be shown that the diagram **(1)** below commutes.

By the universal property of $q_{j^\star(\mathscr{M}')} : F(j^\star(\mathscr{M}')) \to R(j^\star(\mathscr{M}'))$, there is a unique homomorphism of M'-rings $j^R : R(j^\star(\mathscr{M}')) \to R(\mathscr{M}')$ such that the diagram **(2)** below commutes.

For $v : \{x_i : i \in \mathbb{N}\} \to M'$, denote : $H_v^{\mathscr{M}'} : R(\mathscr{M}') \to \mathbb{Z}_2$, the M'-homomorphism $\mathscr{M}'$-coherent; $H_v^{j^\star(\mathscr{M}')} : R(j^\star(\mathscr{M}')) \to \mathbb{Z}_2$, the M'-homomorphism $j^\star(\mathscr{M}')$-coherent both corresponding to v (recall Exercise 27). Then diagram **(3) commutes.**

[18] In fact, $j^\star$ constitutes a contravariant from the category of L'-structures and L'-homomorphisms into the category of L-structures and L-homomorphisms: if $h : \mathscr{M}' \to \mathscr{N}'$ is a L'-homomorphism, then the same map $h : j^\star(\mathscr{M}') \to j^\star(\mathscr{N}')$ is an L-homomorphism.

The commutativity of the external diagram is the content of the equation (EQ) above.



An extension of this theme, that deserves future considerations, is to determine the behavior of the polynomial encoding of FOL under *interpretations of theories*. □

Remark 42 We have already pointed out, in Remark 14, that any Boolean algebra naturally becomes an M-ring. It will be interesting to express (and to explore) the *Boolean-valued models technique* in the polynomial setting, as we did above for the usual Tarskian semantics, i.e., for the Boolean algebra $\mathbb{Z}_2$. In particular, it will be interesting to consider a polynomial version of the following situation:

Let L be a first-order language for each theory and T be a theory (i.e., a subset closed under consequence) in $Form(L)$. Consider S^L the Boolean space whose elements are the maximal consistent theories in $Form(L)$ and denote $S^L_T := \{W \in S^L : T \subseteq W\}$; since S^L_T is a *closed* subset of S^L, then S^L_T is a Boolean (sub)space. Denote $B_T := Clopen(S^L_T)$, the Boolean algebra dual to the Boolean space S^L_T. B_T is not, in general, a complete Boolean algebra, but we can still consider the B_T-Boolean valued models of the theory T.[19] It will be interesting to compare: on one hand, the polynomial version of T given by $\mathscr{R}(L)_T$ (see Definition 26) and, on the other hand,

---

[19]I.e., the class of pairs $\mathscr{M} = (M, [[-]])$, where M is set, $[[-]]$ is a map $(\phi(x_1, \ldots, x_k) \in Form(L)) \mapsto ([[\phi(\bar{x})]] : M^k \longrightarrow B_T)$, satisfying the usual (but conditional, since B_T may not be complete) compatibility requirements of Boolean valued models and, moreover, if $\phi(x_1, \ldots, x_k) \in T$, then $[[\phi(\bar{x})]] = 1_{B_T}$.

the polynomial encodings of the (proper) class of Boolean valued models canonically associated to T, as above. □

Although there is still much work to be done — for instance: to make precise comparisons with other algebraizations of FOL and to develop in further details the theory of M-rings– we believe that the proposal launched in this paper might help to see the relationship between algebra and logic in a new light, overcoming the gulf that separated them in the last centuries, and make the algebraic version of certain logic systems look more natural. This is the case not only with FOL, but with most propositional many-valued and paraconsistent logics. Our aim is that the polynomial method, and certainly their generalizations, might help to reconcile algebra and logic.

References

1. Agudelo, J.C., and W.A. Carnielli. 2011. Polynomial Ring Calculus for modal logics: a new semantics and proof method for modalities. *The Review of Symbolic Logic* 4: 150–170.
2. Béziau, J. -Y. 2007. *Logica Universalis: Towards a general theory of Logic*. Springer.
3. Béziau, J. -Y. 2012. *Universal Logic: an Anthology - From Paul Hertz to Dov Gabbay*. Springer.
4. Blok, W.J. and D. Pigozzi. 1989. *Algebraizable Logics*. Memoirs of the AMS. **396**, American Mathematical Society, Providence, USA.
5. Boole, G. 1847. *The Mathematical Analysis of Logic, Being an Essay Towards a Calculus of Deductive Reasoning*. Cambridge: Macmillan.
6. Bueno-Soler, J., and W.A. Carnielli. 2005. Possible-translations algebraization for paraconsistent logics. *Bulletin of the Section of Logic* 34 (2): 77–92.
7. Burris, S. 2000. *The Laws of Boole's Thought*. Citeseer. https://www.math.uwaterloo.ca/~snburris/htdocs/MYWORKS/PREPRINTS/aboole.pdf.
8. Carnielli, W.A. 2001. A polynomial proof system for Łukasiewicz logics. In *Second Principia International Symposium*, pp. 6–10.
9. Carnielli, W.A. 2005. Polynomial ring calculus for many-valued logics. In *Proceedings of the 35th International Symposium on Multiple-Valued Logic*, IEEE Computer Society, pp. 20–25.
10. Carnielli, W.A. 2005. Polynomial Ring Calculus for Logical Inference. *CLE e-Prints* 5: 1–17.
11. Carnielli, W.A. 2007. Polynomizing: Logic Inference in Polynomial Format and the Legacy of Boole. In *Model-Based Reasoning in Science, Technology, and Medicine*, ed. by L. Magnani, P. Li, vol. 64, pp. 349–364. Springer.
12. Carnielli, W.A. 2010. Formal polynomials, heuristics and proofs in logic. In *Logical Investigations*, ed. by A.S. Karpenko, vol. 16, pp. 280-294. Institute of Philosophy, Russian Academy of Sciences Publisher.
13. Carnielli, W.A., M.E. Coniglio, and J. Marcos. 2007. Logics of Formal Inconsistency. In *Handbook of Philosophical Logic*, ed. by D. Gabbay, F. Guenthner, vol. 14, pp. 15–107. Springer.
14. Carnielli, W.A., and M. Matulovic. 2014. Non-deterministic semantics in polynomial format. *Electronic Notes in Theoretical Computer Science* 305: 19–34.
15. Halmos, P.R. 1962. *Algebraic Logic*. Chelsea Publishing Co.
16. Henkin, L., J.D. Monk, A. Tarski. 1981. *Cylindric Set Algebras and related structures*. Lecture Notes in Mathematics, vol. 883, pp. 1–129. Springer.
17. Henkin, L., J.D. Monk, and A. Tarski. 1985. *Cylindric Algebras: Part II. Studies in Logic and the Foundations of Mathematics*, vol. 115, North-Holland.
18. Matulovic, M. 2013. Proofs in the Pocket: Polynomials as a Universal Method of Proof. Ph.D. Thesis State University of Campinas (Unicamp), Brazil.
19. Moore, G.H. 1988. The emergence of first-order logic. In *History and Philosophy of Modern Mathematics*, ed. by W. Aspray, P. Kitcher, pp. 95–135. University of Minnesota Press.

20. Pinter, C.C.A. 1973. A simple algebra of first order logic. *Notre Dame Journal of Formal Logic* XIV(3): 361–366.
21. Peckhaus, V. 1998. Hugh MacColl and the German algebra of logic. *Nordic Journal of Philosophical Logic* 3 (1): 17–34.
22. Sheffer, H.M. 1913. A set of five independent postulates for Boolean algebras, with application to logical constants. *Transactions of the American Mathematical Society* 14 (4): 481–488.

Plug and Play Negations

Sérgio Marcelino, Carlos Caleiro and Umberto Rivieccio

Abstract We study an array of logics defined on a small set of connectives (including an implication $\to$ and a bottom particle $\bot$) by modularly considering subsets of a set of inference rules that we fix at the start of the game. We provide complete semantics based on possibly non-deterministic logical matrices and complexity upper bounds for the considered logics. As a consequence of the techniques applied, we also obtain completeness results for the negation-only fragments (obtained by defining the negation connective as $\neg p := p \to \bot$, as usual) of all the above-mentioned logics, and analyze their possible paraconsistent character.

1 Introduction

We propose to study the *negations* ($\neg$), defined by the usual abbreviation $\neg p := p \to \bot$ that uses *implication* ($\to$) and *bottom* ($\bot$), on a class of logics modularly defined by considering different subsets of a given fixed set of six Hilbert-style inference rules. We also consider the unary *consistency operator* ($\circ$) stemming from the study of *logics of formal inconsistency* (LFIs) [6], together with one additional new rule for it.

Our strategy will highlight how some key features of negation are inherited from the primitive connectives employed in its definition. Letting the negation connective be defined by means of the abbreviation $\neg p := p \to \bot$ is a fairly standard choice, and allows us to capture not just classical and intuitionistic negation, but also weaker

S. Marcelino (✉) · C. Caleiro
SQIG - Instituto de Telecomunicações, Dep. Matemática – Instituto Superior Técnico, Univ. Lisboa, Lisboa, Portugal
e-mail: smarcel@math.tecnico.ulisboa.pt

C. Caleiro
e-mail: ccal@math.tecnico.ulisboa.pt

U. Rivieccio
Dep. Informática e Matemática Aplicada, Univ. Federal do Rio Grande do Norte, Natal, Brazil
e-mail: urivieccio@dimap.ufrn.br

W. Carnielli and J. Malinowski (eds.), *Contradictions, from Consistency to Inconsistency*, Trends in Logic 47, https://doi.org/10.1007/978-3-319-98797-2_14

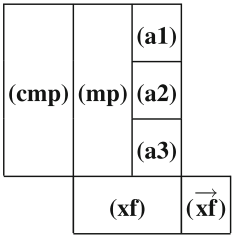


Fig. 1 The playing board

connectives defined from the residuated implications of fuzzy/substructural systems (see [8, 10]). Indeed, we provide a fine-grained analysis showing how certain features of negation depend on properties of the underlying implication connective but also of the bottom nullary connective.

The title of the present paper is meant as a suggestion for the reader to go through it as if he were watching a game being played on a board containing the 7 inference rules depicted in Fig. 1. Each configuration of the board consists of the simultaneous activation of a subset of the rules. Instead of considering all the possible $2^7 = 128$ systems, we shall focus on 12 of the most interesting choices.

We analyze the effect of adding/removing rules from a given configuration, and as an outcome provide simple matrix-based semantics and complexity upper bounds for the logics considered. We shall pay special attention to the resulting negation connectives, and to the way in which the explosiveness/paraconsistency opposition arises in this context. Our analysis will allow us to obtain semantics, and Hilbert style axiomatizations, for each of the negation connectives obtained. We single out 6 distinct negations, 3 of which are paraconsistent. Our choices are summarized in Fig. 4.

From a technical point of view, this paper is an application of recent and fairly general results and techniques aiming at characterizing finite-valuedness in logic [5, 12], the semantics of logics obtained by disjoint fibring and their decidability [4, 13]. Together these results offer us an effective toolkit for the analysis of Hilbert-style systems, especially those obtained by joining calculi for disjoint sets of connectives, providing in particular semantics based on *non-deterministic logical matrices (Nmatrices)* [2, 3, 13]. Without those techniques it would be hard to analyze these logics in a systematic way, despite the familiar look of the inference rules under consideration. Instead, taking advantage of them, we will be able to obtain complete semantics and **coNP**-complexity upper bounds (which, using results from [4], we manage to tighten down to **P** in a few cases) for each of the logics under consideration. We shall also isolate the negation-only fragments of the considered logics, and show that they can all be characterized by a single finite matrix or a single finite non-deterministic matrix where no finite deterministic matrix would be sufficient [2, 3, 5]. We abstain from including in this paper the definitions and results used, referring the reader to the original papers, but still we will illustrate some of the concepts and constructions by means of suitable examples. This choice has the advantage of keeping the plug and play game at the forefront at all times.

We should mention that some of the results obtained here are not new, but our main aim is to illustrate a strategy for analyzing in a modular way similar situations by direct application of general technical tools.

The game shall proceed as follows. In Sect. 2 we lay down the basic board layout. Section 3 is devoted to studying configurations obtained while retaining explosiveness of the bottom connective. Section 4 parallels the preceding one, but focuses on configurations of the game obtained by getting rid of explosiveness of the bottom. We conclude in Sect. 5 with a brief discussion concerning future work.

2 Basics of the Game

The setup of our plug and play game is quite easy. We will consider logics defined from two primitive connectives only: the binary $\to$ implication and the nullary connective $\bot$. Negation is defined in all cases as $\neg p := p \to \bot$. Later on we will further consider a unary consistency-type operator.

We assume that the reader is familiar with consequence relations, logical matrices (deterministic and non-deterministic) and many-valued logics; for all unexplained terminology we refer the reader to [3, 9, 17].

We will use CL and IL to denote, respectively, classical and intuitionistic propositional logic. Given any subset S of the connectives of a logic L, we will denote by L^{S} its corresponding S-fragment. Since $\mathsf{CL}^{\bot} = \mathsf{IL}^{\bot}$, we will dub $\bot$, for *bottom*, the bottom-only fragment of both logics (which is actually the same for every non-trivial logic with an explosive nullary connective).

2.1 The Game Board

Our game board, depicted in Fig. 1, consists of the 7 rules defined in Fig. 2, which are all very simple and mostly familiar. The core, consisting of *modus ponens* **(mp)** plus the axioms **(a1)**-**(a3)**, allows us to define the implication-only fragments of classical or intuitionistic logic, depending on whether we (un)plug axiom **(a3)**, as well as the implication fragments of other substructural logics. The board is extended at the bottom with 2 rules concerning $\bot$: the *ex falso* rule **(xf)** and the axiom $(\overrightarrow{\mathbf{xf}})$ which involves the implication too. Notice that, in the presence of **(mp)**, $(\overrightarrow{\mathbf{xf}})$ subsumes **(xf)**. Finally, we extend the board to the left with one additional rule **(cmp)** which involves the consistency operator, and can be viewed as a form of *cautious modus ponens*.

$$\frac{p \quad p \to q}{q}\ \textbf{(mp)} \qquad \frac{\circ p \quad p \quad p \to q}{q}\ \textbf{(cmp)}$$

$$\frac{}{p \to (q \to p)}\ \textbf{(a1)} \qquad \frac{}{(p \to (q \to r)) \to ((p \to q) \to (p \to r))}\ \textbf{(a2)}$$

$$\frac{}{((p \to q) \to p) \to p}\ \textbf{(a3)}$$

$$\frac{\bot}{p}\ \textbf{(xf)} \qquad \frac{}{\bot \to p}\ (\overrightarrow{\textbf{xf}})$$

Fig. 2 Inference rules

2.2 Classical and Intuitionistic Logics

In the initial configuration of the game board only **(cmp)** is switched off, and the rules **(mp) (a1) (a2) (a3) (xf)** ($\overrightarrow{\textbf{xf}}$) are all switched on. The logic thus defined is well known to be the $\{\to, \bot\}$-fragment of CL, which we will also call CL.

A complete semantics for CL is provided by the standard $\{0, 1\}$-valued Boolean matrix $\mathbb{CL}$ (with 1 as designated value) given by the tables:

$\to$	0	1
0	1	1
1	0	1

$\bot$
0

The $\{\neg\}$-fragment of CL obtained by defining $\neg p := p \to \bot$ is classical negation ($\mathsf{CL}^{\neg}$) which can be axiomatized by the rules of *double negation introduction*, *double negation elimination*, and *ex contradictione*:

$$\frac{p}{\neg\neg p}\ \textbf{(di)} \qquad \frac{\neg\neg p}{p}\ \textbf{(de)} \qquad \frac{p \quad \neg p}{q}\ \textbf{(xc)}$$

Semantically, classical negation is characterized by the reduct $\mathbb{CL}^{\neg}$ of the 2-valued Boolean matrix:

	$\neg$
0	1
1	0

The game goes on by switching on or off some of the rules, thus changing the configuration of the board. An obvious first move is to switch Peirce's axiom **(a3)** off. The resulting logic **(mp) (a1) (a2) (xf)** ($\overrightarrow{\textbf{xf}}$) is the $\{\to, \bot\}$-fragment of intuitionistic logic, which we also[1] dub IL. This calculus is complete with respect to the class of all matrices $\langle \mathbf{A}, \{1\}\rangle$ where $\mathbf{A}$ is *Hilbert algebra* with minimum element, i.e. $\mathbf{A}$ is

[1]Note that in CL this is quite natural as all the other classical connectives are definable from $\to$ and $\bot$. In IL this is not the case, as for example neither conjunction nor disjunction can be defined from $\to$ and $\bot$. Still, for neatness of presentation, we commit this abuse of notation also here.

the $\{\to, \bot\}$-subreduct of a Heyting algebra, with $\bot$ interpreted as the minimum and having 1 as maximum.[2]

The $\{\neg\}$-fragment of IL, denoted $\mathsf{IL}_{\neg}$, can be axiomatized by weakening the rule of double negation elimination **(de)** into **(wde)**:

$$\frac{p}{\neg\neg p}\ \textbf{(di)} \qquad \frac{\neg\neg\neg p}{\neg p}\ \textbf{(wde)} \qquad \frac{p \quad \neg p}{q}\ \textbf{(xc)}$$

Also, $\mathsf{IL}^{\neg}$ is is characterized by the $\{0, \frac{1}{2}, 1\}$-valued matrix[3] $\mathbb{G}_3^{\neg}$ with 1 designated given by the table

	$\neg$
1	0
$\frac{1}{2}$	0
0	1

Another natural move in the game would be to switch **(xf)** off. However, since **(xf)** is subsumed by the joint efforts of **(mp)** and $(\overrightarrow{\textbf{xf}})$, the logic would simply not change, both in the classical and in the intuitionistic case. A more interesting move is to switch $(\overrightarrow{\textbf{xf}})$ off, which we are going to do in the next section.

3 *Ex falso* Logics

We now switch the $(\overrightarrow{\textbf{xf}})$ rule off for good.

3.1 Intuitionistic Negation

We start with the following configuration: **(mp) (a1) (a2) (a3) (xf)**. The resulting logic is denoted by $\mathsf{CL}^{\to} \bullet \bot$ and corresponds to the *disjoint fibring* [13] of the $\{\to\}$-fragment with the $\{\bot\}$-fragment of classical logic. The following proposition shows that $\mathsf{CL}^{\to} \bullet \bot$ is four-valued.

Proposition 1 *$CL^{\to} \bullet \bot$ is characterized by the matrix $\mathbb{M}_4$ with $D = \{11\}$ given by the tables*

[2]The class of Hilbert algebras [11] is the algebraic semantics of the $\{\to\}$-fragment of intuitionistic logic. Hilbert algebras, also called *(positive) implication(al) algebras* [1, 15], correspond to $\{\to\}$-subreducts of Heyting algebras and thus have a definable natural order which has 1 as maximum element but need not have a minimum.

[3]Which is the $\{\neg\}$-reduct of $\mathbb{G}_3$, corresponding to three-valued Gödel logic, see [9].

$\to$	00	01	10	11
00	11	11	11	11
01	10	11	10	11
10	01	01	11	11
11	00	01	10	11

$\bot$
10

(notice that $\bot$ is not interpreted as 00 *as one might expect).*

Proof $\mathsf{CL}^{\to}$-soundness of the matrix follows from the fact that the implication is defined as in the product Boolean algebra **2** $\times$ **2**, and any matrix that interprets $\bot$ outside of D is $\mathbf{\bot}$-sound.

As to completeness, note that in the logic $\mathsf{CL}^{\to} \bullet \bot$ we have that $\Gamma \vdash \varphi$ if and only if $\Gamma \vdash \varphi$, or else $\Gamma \vdash \bot$, in both cases using only the rules **(mp)** and **(a1)-(a3)** of $\mathsf{CL}^{\to}$. Therefore, assuming that $\Gamma \nvdash \varphi$, and taking $\bot$ as an additional atomic particle, we know that there are valuations v_1, v_2 over the implication-reduct of the 2-valued Boolean matrix such that $v_1(\Gamma) = v_2(\Gamma) = \{1\}$ and $v_1(\varphi) = v_2(\bot) = 0$. Hence, we can define a valuation over the 4-valued matrix as follows:

$$v(\psi) = \begin{cases} 1v_1(\psi) & \text{if } v_1(\bot) = 0 \\ v_1(\psi)v_2(\psi) & \text{if } v_1(\bot) = 1 \end{cases}.$$

When $v_1(\bot) = 0$ it is clear that v is such that, for any formula ψ, $v(\psi) = 1v_1(\psi) =$ 11 if and only if $v_1(\psi) = 1$ (of course, $v(\bot) = 1v_1(\bot) = 10$). Thus, $v(\Gamma) = \{11\}$ and $v(\varphi) = 1v_1(\varphi) = 10 \neq 11$.

When $v_1(\bot) = 1$, similarly, v is such that, for any formula ψ, $v(\psi) = v_1(\psi)v_2(\psi)$ $= 11$ if and only if $v_1(\psi) = v_2(\psi) = 1$ (of course, $v(\bot) = v_1(\bot)v_2(\bot) = 10$). Thus, again, $v(\Gamma) = \{11\}$ and $v(\varphi) = v_1(\varphi)v_2(\varphi) = 0v_2(\varphi) \neq 11$.

Note that we implicitly use the fact that the valuation that sends every formula to 1 is a sound model for $\to$. Indeed, our construction in this proof would apply not just to classical $\to$, but *mutatis mutandis* to any truth-preserving connective. □

The matrix $\mathbb{M}_4^{\neg}$ for negation in $\mathsf{CL}^{\to} \bullet \bot$ that we can derive from the above matrix is given by the table

	$\neg$
00	11
01	10
10	11
11	10

In fact, $\mathbb{M}_4^{\neg}$ defines the same logic as $\mathbb{G}_3^{\neg}$. To see this, notice that, since the four-valued negation is sound with respect to the inference rules **(di) (wde) (xc)** of $\mathsf{IL}^{\neg}$, we know that the $\{\neg\}$-fragment of $\mathsf{CL}^{\to} \bullet \bot$ is at least as strong as intuitionistic negation. To see that the two are in fact the same, it is then enough to observe that $\mathbb{M}_4^{\neg}$ is the quotient of $\mathbb{M}_4$ by the matrix congruence that identifies (only) the elements 00 and 10.

It might come as a surprise that the negation of $\mathsf{CL}^{\rightarrow} \bullet \bot$ coincides with the intuitionistic one; but given this, one easily sees that further removing **(a3)** from $\mathsf{CL}^{\rightarrow} \bullet \bot$ does not affect its negation fragment. So let us consider the configuration given by the rules **(mp) (a1) (a2) (xf)**, which gives us the logic $\mathsf{IL}^{\rightarrow} \bullet \bot$. This is the disjoint fibring of the $\{\rightarrow\}$-fragment with the $\{\bot\}$-fragment of intuitionistic logic.

In order to check that the negation fragment of $\mathsf{IL}^{\rightarrow} \bullet \bot$ also coincides with $\mathsf{IL}^{\neg}$, we shall prove a stronger result, namely that this happens for a larger class logics: in fact, for all logics between $\mathsf{IL}^{\rightarrow} \bullet \bot$ and its strengthening with **(a3)** (that is, $\mathsf{CL}^{\rightarrow} \bullet \bot$) and also for all the logics between $\mathsf{IL}^{\rightarrow} \bullet \bot$ and its strengthening with $(\overrightarrow{\mathbf{xf}})$ (that is, IL).

Denote by $[\mathsf{L}_1, \mathsf{L}_2]$ the class of all logics (in the $\{\rightarrow, \bot\}$-language) that are stronger than L_1 and weaker than L_2.

Proposition 2 *For every logic* $\mathsf{L} \in [\mathsf{IL}^{\rightarrow} \bullet \bot, \mathsf{CL}^{\rightarrow} \bullet \bot] \cup [\mathsf{IL}^{\rightarrow} \bullet \bot, \mathsf{IL}]$*, defining as usual* $\neg p := p \rightarrow \bot$*, we have* $\mathsf{L}^{\neg} = \mathsf{IL}^{\neg}$.

Proof As already noted in [14], $\mathsf{IL}^{\rightarrow} \bullet \bot$ is given by the class of all matrices $\langle \mathbf{A}, \{1\} \rangle$ where $\mathbf{A}$ is a Hilbert algebra with 1 as maximum and $\bot$ is interpreted as any non-designated element (not necessarily the minimum, which may not even exist).

Any Hilbert algebra satisfies, for all $a, b \in \mathbf{A}$, the following (in)equalities:

$$
\begin{aligned}
1 \rightarrow a &= a \\
a &\leq (a \rightarrow b) \rightarrow b \\
((a \rightarrow b) \rightarrow b) \rightarrow b &\leq a \rightarrow b.
\end{aligned}
$$

These easily imply that every matrix $\langle \mathbf{A}, \{1\} \rangle$ is sound with respect to the rules in $R_{\mathsf{IL}^{\neg}}$. That is, it is not possible to have both $v(\neg\varphi) := v(\varphi) \rightarrow v(\bot) = 1$ and $v(\varphi) = 1$, that $v(\varphi) = 1$ implies $(v(\varphi) \rightarrow v(\bot)) \rightarrow v(\bot) = 1$, and that $v(\neg\neg\neg\varphi) = ((v(\varphi) \rightarrow v(\bot)) \rightarrow v(\bot)) \rightarrow v(\bot) = 1$ implies $v(\neg\varphi) = v(\varphi) \rightarrow v(\bot) = 1$.

Since $\mathsf{IL}^{\rightarrow} \bullet \bot$ is (strictly) weaker that IL it must also define the intuitionistic negation (fragment). Since we have already seen that intuitionistic negation is also the (defined) negation fragment of $\mathsf{CL}^{\rightarrow} \bullet \bot$, the result follows. □

Although all the logics mentioned in Proposition 2 share the same negation fragment, it is worth noting that they can differ significantly over the full language. The class $[\mathsf{IL}^{\rightarrow} \bullet \bot, \mathsf{CL}^{\rightarrow} \bullet \bot] \cup [\mathsf{IL}^{\rightarrow} \bullet \bot, \mathsf{IL}]$ obviously contains logics that are not disjoint fibrings, e.g. all the $\{\rightarrow, \bot\}$-fragments of Gödel-Dummett finite and infinite-valued logics [10].

From a complexity point of view, we may also observe that while for instance $\mathsf{CL}^{\rightarrow} \bullet \bot$ is **coNP**-complete (like classical logic), $\mathsf{IL}^{\rightarrow} \bullet \bot$ and IL are **PSPACE**-complete [16]. It is also easy to see that in the class $[\mathsf{IL}^{\rightarrow} \bullet \bot, \mathsf{CL}^{\rightarrow} \bullet \bot] \cup [\mathsf{IL}^{\rightarrow} \bullet \bot, \mathsf{IL}]$ there are undecidable logics as well.

3.2 Isolating ex Contradictione

When we further drop **(a2)** we are left with the rules **(mp) (a1) (xf)** which characterize the disjoint fibring $\mathsf{MP}_1 \bullet \bot$, where MP_1 denotes the logic given by the rules **(mp) (a1)**.

Proposition 3 *MP_1 cannot be characterized by any finite Nmatrix.*

Proof It is enough to show that MP_1 is not finitely-determined in the sense of [5, Definition 3.1]. Although [5] considers only deterministic matrices, it is easy to see that a logic given by a finite Nmatrix must also be finitely-determined.

Define, for every natural n,

$$\begin{aligned} \gamma(p_1, \ldots, p_n) &= p_1 \to (p_2 \to \ldots (p_{n-1} \to p_n) \ldots) \\ \psi(p_1, \ldots, p_{n+1}) &= \gamma(p_1, \ldots, p_n) \to (p_1 \to \gamma(p_2, \ldots, p_{n+1})) \\ \mathbf{n} &= \{p_1, \ldots, p_n\} \\ \Gamma &= \{\psi(p_1, \ldots, p_{n+1})^\sigma \mid \sigma : \mathbf{n+1} \to \mathbf{n+1}, \sigma \neq Id_{\mathbf{n+1}}\} \end{aligned}$$

It is not hard to check that $\Gamma \not\vdash \psi(p_1, \ldots, p_{n+1})$. However, $\Gamma^\tau \not\vdash \psi(p_1, \ldots, p_{n+1})^\tau$ for every $\tau : \mathbf{n+1} \to \mathbf{n}$ as $\psi(p_1, \ldots, p_{n+1})^\tau \in \Gamma^\tau$. □

We are now going to introduce a (necessarily infinite) Nmatrix $\mathbb{MP}_1$ that does characterize MP_1. Let **Fm** denote the set of formulas of MP_1 built from a denumerable set of propositional variables P. The universe of our matrix is $A = \mathbf{Fm} \times \{0, 1\}$, with designated elements $D = \mathbf{Fm} \times \{1\}$. The implication is given by

$$\begin{aligned} (\varphi, a) \to (\psi, b) = \{(\varphi \to \psi, c) :\ & \\ & \text{if } a = 1 \text{ and } b = 0 \text{ then } c = 0, && (1) \\ & \text{if } b = 1 \text{ then } c = 1, && (2) \\ & \text{if } \psi = \psi' \to \varphi \text{ and } a = 0 \text{ then } c = 1\}. && (3) \end{aligned}$$

Proposition 4 *MP_1 is characterized by the saturated Nmatrix $\mathbb{MP}_1$.*

Proof Note that for every formula and valuation over $\mathbb{MP}_1$ we have that $\pi_1(v(\varphi)) = \varphi^{\sigma_v}$ where $\sigma_v \colon P \to \mathbf{Fm}$ is given by $\sigma_v(p) = \pi_1(v(p))$.

Let us check that MP_1 is sound with respect to $\mathbb{MP}_1$. **(mp)**-soundness follows immediately from (1). To see that it is also **(a1)**-sound, we just have to consider two cases, either $v(\varphi) = (\varphi^{\sigma_v}, 1)$ or $v(\varphi) = (\varphi^{\sigma_v}, 0)$. In both cases we obtain that $v(\varphi \to (\psi \to \varphi)) = ((\varphi \to (\psi \to \varphi))^{\sigma_v}, 1)$, either by invoking (2), or by invoking (twice) (3).

To show that $\mathbb{MP}_1$ is saturated and characterizes MP_1, it is enough to check that for every MP_1-theory we have that the map $v_T \colon \mathbf{Fm} \to A$ defined as

$$v_T(\varphi) = \begin{cases} (\varphi, 1) & \text{if } \varphi \in T, \\ (\varphi, 0) & \text{otherwise,} \end{cases}$$

is a valuation over $\mathbb{MP}_1$. We leave the straightforward details to the reader. □

By [13, Theorem 1], we have then that $\mathsf{MP}_1 \bullet \bot$ is characterized by the Nmatrix $\mathbb{MP}_1 \star \mathbb{B}\mathrm{ot}$, that is, the *strict product* (as defined in [13]) of the matrices $\mathbb{MP}_1$ and $\mathbb{B}\mathrm{ot}$, the latter denoting just the $\{0, 1\}$-valued matrix over the language $\{\bot\}$ that interprets $\bot$ as 0. This Nmatrix is equivalently presented by just extending $\mathbb{MP}_1$ with a non-deterministic $\bot$ interpreted outside the set of designated values, that is, over $\mathbf{Fm} \times \{0\}$. It is not hard to check that every formula ψ in the negation only fragment is of (unravels to) the form $\psi = (((p \to \bot) \ldots \to \bot) \to \bot$ for some variable p, therefore is never an instance of **(a1)**, and thus (3) of Proposition 4 never applies. This implies that for each choice of $\bot = (\psi, 0)$ we obtain an Nmatrix equivalent to $\mathbb{MP}_1^{\neg}$ given by the following table with designated 1:

	$\neg$
0	$\{0, 1\}$
1	$\{0\}$.

The resulting logic is therefore characterized by the above Nmatrix and is axiomatized by the single rule

$$\frac{p \quad \neg p}{q}\ \textbf{(xc)}$$

which is the principle of explosion known in the literature as *ex contradictione quodlibet*. Logics which do not satisfy **(xc)** are usually called *paraconsistent*. We shall call a negation satisfying only **(xc)**, *explosion-only negation*.

We now remove also **(a1)**, thus being left with **(mp)** and **(xf)**. This configuration defines the disjoint fibring $\mathsf{MP} \bullet \bot$. The $\{\to\}$-factor in this fibring is the logic of modus ponens MP, which is not complete with respect to any finite matrix,[4] but is characterized by the Nmatrix $\mathbb{MP}$ (where 1 is the only designated element) given by the table:

$\to$	0	1
0	$\{0, 1\}$	$\{0, 1\}$
1	$\{0\}$	$\{0, 1\}$

It is not hard to show that $\mathbb{MP}$ is saturated, hence we could, as before, obtain a table for the resulting negation (this time over $\mathbb{MP} \star \mathbb{B}\mathrm{ot}$) and conclude that it still defines the same negation as $\mathbb{MP}_1 \star \mathbb{B}\mathrm{ot}$ (we actually obtain the same table). It is however enough to observe that **(xc)** is derivable in $\mathsf{MP} \bullet \bot$, and since $\mathsf{MP} \bullet \bot$ is weaker than $\mathsf{MP}_1 \bullet \bot$, it cannot have a stronger negation: hence the negations of the two logics must coincide.

[4]This easily follows from [5] and the fact that MP is weaker than IL.

3.3 Playing Ball

If we drop **(mp)** and consider configurations of the board involving only the rules **(a1) (a2) (a3) (xf)**, then all the resulting negations coincide and are given by the empty set of rules. We shall call this negation logic the *free negation*. Obviously, over larger languages interactions may appear, but only as theorems and not as proper inference rules. For example, $\bot \to \neg\bot$ is a theorem whenever **(a1)** is on. Free negation is paraconsistent in the sense that **(xc)** is not valid, but it is not very interesting because such paraconsistency is due to the fact that the logic is extremely weak.

As we shall see in the next section, dropping **(xf)** (with ($\overrightarrow{\mathbf{xf}}$) still off) leads to paraconsistent logics that retain some reasoning power. In the remainder of this section, we shall however focus on an alternative way of achieving paraconsistency. That is, we replace **(mp)** by its *cautious* version **(cmp)** displayed in Fig. 2, introducing the consistency operator $\circ$. In this way the rule that would produce explosion can only be applied if the antecedent of the implication is not only true but also "consistent".

Consider then the configuration having only **(cmp)** and **(xf)** on. The corresponding logic is $\mathsf{MP}_\circ \bullet \bot$. The next proposition states that its $\{\to, \circ\}$-fragment $\mathsf{MP}_\circ$ can be characterized by the three-element Nmatrix $\mathbb{MP}_\circ$ with operations defined as below, with values $A = \{0, \frac{1}{2}, 1\}$ and designated elements $D = \{\frac{1}{2}, 1\}$.

$\to$	0	$\frac{1}{2}$	1
0	A	A	A
$\frac{1}{2}$	A	A	A
1	$\{0\}$	A	A

	$\circ$
0	A
$\frac{1}{2}$	$\{0\}$
1	A

Proposition 5 *$\mathsf{MP}_\circ$ is characterized by the saturated Nmatrix $\mathbb{MP}_\circ$.*

Proof Let us verify that $\mathbb{MP}_\circ$ is complete and saturated. Soundness is routine checking. We prove completeness and saturation at once by defining, for every theory T closed under **(cmp)**, a valuation v as follows: for any formula ψ,

$$v(\psi) = \begin{cases} 1 & \text{if } \{\psi, \circ\psi\} \subseteq T \\ \frac{1}{2} & \text{if } \psi \in T \text{ and } \circ\,\psi \notin T \\ 0 & \text{if } \psi \notin T. \end{cases}$$

We show by induction that v is a valuation over the above defined matrix. For atomic formulas there is nothing to check. Let ψ be a compound formula, and suppose $\psi = \circ\psi_1$. Assume $\psi \in T$, and let us show that $v(\psi) \neq 0$ respects the structure of the Nmatrix. Now $v(\psi) = 0$ would be forced only by $v(\psi_1) = \frac{1}{2}$, which would imply (by the induction hypothesis) that $\circ\psi_1 \notin T$, contradicting our assumption. Conversely, if $\psi \notin T$, then regardless of $v(\psi_1)$, evaluating $v(\psi) = 0$ will respect the structure of the Nmatrix. Let now $\psi = \psi_1 \to \psi_2$. Assume $\psi \in T$. Now $v(\psi) = 0$

would be forced only by $v(\psi_1) = 1$ and $v(\psi_2) = 0$. By induction hypothesis, this would mean that $\psi_1, \circ\psi_1 \in T$ and $\psi_2 \notin T$, which is precisely the case that the rule of $\mathsf{MP}_\circ$ forbids. Conversely, assuming $\psi \notin T$, then regardless of $v(\psi_1)$ and $v(\psi_2)$, evaluating $v(\psi) = 0$ will respect the structure of the Nmatrix. $\square$

By [13, Theorem 1], we then have that $\mathsf{MP}_\circ \bullet \bot$ is characterized by the strict product $\mathbb{MP}_\circ \star \mathbb{B}\mathbb{o}\mathbb{t}$ with underlying set $A' = \{00, \frac{1}{2}1, 11\}$, designated elements $D = \{\frac{1}{2}1, 11\}$ and operations given by the tables below.

$\to$	00	$\frac{1}{2}1$	11
00	A'	A'	A'
$\frac{1}{2}1$	A'	A'	A'
11	$\{00\}$	A'	A'

	$\circ$
00	A'
$\frac{1}{2}1$	$\{00\}$
11	A'

$\bot$
00

The defined negation is given by the table

$\neg$	
00	A'
$\frac{1}{2}1$	A'
11	$\{00\}$

Since $\frac{1}{2}1$ is designated, one easily obtains that the negation fragment of $\mathsf{MP}_\circ \bullet \bot$ coincides with the free negation as happened with $\mathsf{MP} \bullet \bot$, and it is axiomatized by the empty set of rules. The $\{\neg, \circ\}$-fragment is more interesting, and is given by the single rule

$$\frac{\circ p \quad p \quad \neg p}{q}\ \textbf{(gxc)}$$

which is the "Finite Gentle Principle of Explosion" considered for example in [6, Section 3.2].

4 Dropping *ex Falso*

After exploring various outcomes of the board configurations with **(xf)** on, we study the effect of turning **(xf)** off on each of the previously considered configurations. On the semantical side, this amounts to allowing $\bot$ to be interpreted as a designated element. This fact has already been observed in [14], where the free $\bot$ is called *botop* and denoted $\mathbb{I}$. Let then $\mathbb{I}$ be the logic of the free nullary connective (for which we keep the same name $\bot$), that is, the logic axiomatized by the empty set of rules. Clearly, $\mathbb{I}$ is characterized by $\mathbb{B}\mathbb{o}\mathbb{t}\mathbb{o}\mathbb{p}$, the $\{0, 1\}$-valued Nmatrix where $\bot$ is interpreted (freely) as $\{0, 1\}$. So all the logics considered in this section result from expanding the $\bot$-free fragments of any logic L examined before with a free $\bot$. This corresponds to the disjoint fibring $\mathsf{L}^{\to} \bullet \mathbb{I}$. This construction was shown in [13, Proposition 1] to be captured semantically by adding a free operator to the

Nmatrices characterizing the initial logic. Expanding an Nmatrix $\mathbb{M}$ with a free nullary connective corresponds to the strict product $\mathbb{M} \star \mathbb{B}\mathbb{o}\mathbb{t}\mathbb{o}\mathbb{p}$. When the matrix we started with is deterministic, this idea can be mimicked deterministically (for a nullary connective such as $\bot$) by considering one matrix for each possible choice of value for $\bot$.

4.1 Johansson's Minimal Negation

Here we consider various configurations of the board, all defining the $\{\neg\}$-fragment of Johansson's minimal logic, which we dub *Johansson's minimal negation*. As we shall see, this logic is axiomatized by the rules of introduction and elimination of intuitionistic negation **(di)** and **(wde)**, and **(wxc)**, a gentler version of the explosion rule **(xc)**:

$$\frac{p}{\neg\neg p}\ \textbf{(di)} \qquad \frac{\neg\neg\neg p}{\neg p}\ \textbf{(wde)} \qquad \frac{p \quad \neg p}{\neg q}\ \textbf{(wxc)}$$

We can therefore say that this negation is a paraconsistent intuitionistic negation. It is relevant to note that we also conclude that this negation is definable in a $\{\to, \bot\}$-Nmatrix with four values.

Let us go back to the configuration having all rules for classical implication **(mp)** **(a1)** **(a2)** **(a3)** on, but all rules involving the other connectives off. The $\{\to, \bot\}$-fragment of the resulting logic corresponds to $\mathsf{CL}^{\to} \bullet \mathbb{I}$. As mentioned above, this logic is characterized by the expansion of the matrix of classical implication with a free nullary connective $\bot$. Or, deterministically, this can be accomplished with two $\{0, 1\}$-valued matrices, both having a classical implication, and such that $\bot$ takes, in each matrix, one of the two possible values. The resulting negation is given by the matrices $\mathbb{CL}^{\neg}$ and $\mathbb{M}_2^{\neg}$:

	$\neg$
0	1
1	0

	$\neg$
0	1
1	1

Valuations over $\mathbb{CL}^{\neg}$ are classical, while over $\mathbb{M}_2^{\neg}$ valuations can only refute propositional variables because all negated formulas are always designated. The rule **(de)** is thus not sound in $\mathbb{M}_2^{\neg}$, but its intuitionistic version **(wde)** is. It is straightforward to check that the above matrices are sound with respect to the rules **(di)**, **(wde)** and **(wxc)**.

To see that these rules actually capture the negation of $\mathsf{CL}^{\to} \bullet \mathbb{I}$, let T be the closure of Γ under **(di)**, **(wde)** and **(wxc)**, and let $\varphi \notin T$. It is easy to see that if $\varphi = p$ (is a variable), the map v defined, for all propositions ψ, by

$$v(\psi) = \begin{cases} 1 & \text{if } \psi \neq \varphi \\ 0 & \text{otherwise,} \end{cases}$$

is a valuation over $\mathbb{M}_2$, and that $v(\Gamma) = \{1\}$ and $v(\varphi) = 0$. Otherwise, $\varphi = \neg^n p$ for some $n > 0$, and therefore, from **(wxc)** we know there is no ψ such that $\psi \in T$ and $\neg\psi \in T$. Consider the valuation over $\mathbb{CL}^{\neg}$, v given by

$$v(q) = \begin{cases} 1 & \text{if } \neg q \notin T \\ 0 & \text{otherwise.} \end{cases}$$

From **(di)** and **(wde)** we conclude that $v(\Gamma) = \{1\}$ and $v(\varphi) = 0$.

As **(di)** and **(wde)** are the rules of introduction and elimination of intuitionistic negation and **(wxc)** is a gentler version of the explosion rule **(xc)**, we may call this negation a paraconsistent intuitionistic negation. Let us show that it coincides with the $\{\neg\}$-fragment of Johansson's minimal logic. The $\{\to, \bot\}$-fragment of Johansson's minimal logic corresponds to $\mathsf{IL}^{\to} \bullet \mathbb{I}$, as observed in [14]. This logic is given by the rules left after dropping **(a3)** from the current configuration, that is, **(mp) (a1) (a2)**. Again, we know that $\mathsf{IL}^{\to} \bullet \mathbb{I}$, is characterized by the class of Hilbert algebras in which $\bot$ can be interpreted as any element. However, when dealing with intuitionistic negation we observed that valuations of the defined negation over Hilbert algebras where $\bot$ is not the maximum element are captured by valuations over $\mathbb{G}_3^{\neg}$. It is also easy to see that if $\bot$ is the maximum element then every negated formula is designated. We can therefore conclude that Johansson's minimal negation is given by the two matrices $\mathbb{G}_3^{\neg}$ and $\mathbb{M}_2^{\neg}$ introduced earlier.

To see that both pairs of matrices define the same logic, it is enough to note that $\mathbb{G}_3^{\neg}$ is sound for the rules **(di) (wde) (wxc)** and, since $\mathsf{IL}^{\to} \bullet \mathbb{I}$ is (strictly) weaker than $\mathsf{CL}^{\to} \bullet \mathbb{I}$, it cannot define a stronger negation.

4.2 Isolating Weak ex Contradictione

We now further remove **(a2)**, being left with only **(mp)** and **(a1)**, which gives us the logic $\mathsf{MP}_1 \bullet \mathbb{I}$. As mentioned above, $\mathsf{MP}_1 \bullet \mathbb{I}$ is characterizable by freely interpreting $\bot$ over the matrix $\mathbb{MP}_1$. Note that when $\bot$ is non-designated we obtain the Nmatrix $\mathbb{MP}_1^{\neg}$ as discussed in Sect. 3.2, and when it is a designated element we obtain a Nmatrix equivalent to $\mathbb{M}_2^{\neg}$ as every negation becomes designated. Hence, the negation fragment of $\mathsf{MP}_1 \bullet \mathbb{I}$ is characterized by the two Nmatrices $\mathbb{MP}_1^{\neg}$ and $\mathbb{M}_2^{\neg}$. This negation is axiomatized by the weak explosion rule alone:

$$\frac{p \quad \neg p}{\neg q}\,\textbf{(wxc)}$$

$$\frac{p}{\neg\neg p}\,\textbf{(di)} \qquad \frac{\neg\neg p}{p}\,\textbf{(de)} \qquad \frac{\neg\neg\neg p}{\neg p}\,\textbf{(wde)}$$

$$\frac{p \quad \neg p}{q}\,\textbf{(xc)} \qquad \frac{p \quad \neg p}{\neg q}\,\textbf{(wxc)} \qquad \frac{\circ p \quad p \quad \neg p}{q}\,\textbf{(gxc)}$$

Fig. 3 Negation rules

4.3 Free Negation Again

The same strategy can of course be applied when we further remove **(a1)** and are left with just **(mp)**. The resulting logic $\mathsf{MP}\bullet\,\mathbb{I}$ is characterizable by freely interpreting $\bot$ over $\mathbb{MP}$. The negation of $\mathsf{MP}\bullet\,\mathbb{I}$ is characterized by the two Nmatrices given by the tables corresponding to each of the $\mathbb{MP}$ columns:

	$\neg$
0	{0, 1}
1	{0}

	$\neg$
0	{0, 1}
1	{0, 1}

We shall denote by $\mathbb{F}^{\neg}$ the matrix given by the second table. It is easy to see that $\mathbb{F}^{\neg}$ defines the free negation introduced in Sect. 3.3, and that any class of matrices containing $\mathbb{F}^{\neg}$ will define the same logic. And so, where before (in Sect. 3.2) we had obtained a negation captured simply by the rule of explosion, here we obtain the negation characterized by the empty set of rules.

4.4 Playing Ball Again

The logic $\mathbb{MP}_{\circ}\bullet\,\mathbb{I}$ is given by freely interpreting $\bot$ as any element in $A = \{0, \frac{1}{2}, 1\}$ over the matrix $\mathbb{MP}_{\circ}$ with operations defined as below and designated elements $D = \{\frac{1}{2}, 1\}$.

$\to$	0	$\frac{1}{2}$	1
0	A	A	A
$\frac{1}{2}$	A	A	A
1	{0}	A	A

	$\circ$
0	A
$\frac{1}{2}$	{0}
1	A

The resulting negation is given by three matrices, one for each column (i.e. each choice for the element interpreting $\bot$) of the implication table. The rule **(gxc)** is not sound in $\mathbb{MP}_{\circ} \star \mathbb{B}\mathrm{otop}$, and in fact the $\{\neg, \circ\}$-fragment defined over $\mathbb{MP}_{\circ}\bullet\,\mathbb{I}$ is axiomatized by the empty set of rules. To see this, it is sufficient to observe that the free matrix (where $\neg a = \circ a = A$ for each element in A) is saturated with respect to the logic defined by the empty set of rules.

Logic	$\{\to, \bot, \circ\}$-Rules	$\{\neg, \circ\}$-Rules	Matrices	$\{\neg\}$-Matrices	Parac.
CL	**(mp) (a1) (a2) (a3) (xf)** $(\overrightarrow{\mathbf{xf}})$	**(xc) (di) (de)**	$\mathbb{CL}$	$\mathbb{CL}^{\neg}$	N
IL	**(mp) (a1) (a2) (xf)** $(\overrightarrow{\mathbf{xf}})$	**(xc) (di) (wde)**	$*$	$\mathbb{G}_3^{\neg}$	N
$\mathsf{CL}^{\to} \bullet \bot$	**(mp) (a1) (a2) (a3) (xf)**	**(xc) (di) (wde)**	$\mathbb{M}_4$	$\mathbb{G}_3^{\neg}$	N
$\mathsf{IL}^{\to} \bullet \bot$	**(mp) (a1) (a2) (xf)**	**(xc) (di) (wde)**	$**$	$\mathbb{G}_3^{\neg}$	N
$\mathsf{MP}_1 \bullet \bot$	**(mp) (a1) (xf)**	**(xc)**	$\mathbb{MP}_1^{\neg} \star \mathbb{B}\mathbb{o}\mathbb{t}$	$\mathbb{MP}_1^{\neg}$	N
$\mathsf{MP} \bullet \bot$	**(mp) (xf)**	**(xc)**	$\mathbb{MP} \star \mathbb{B}\mathbb{o}\mathbb{t}$	$\mathbb{MP}_1^{\neg}$	N
$\mathsf{CL}^{\to} \bullet \top\!\!\!\bot$	**(mp) (a1) (a2) (a3)**	**(wxc) (di) (de)**	$\mathbb{CL}^{\to} \star \mathbb{B}\mathbb{o}\mathbb{t}\mathbb{o}\mathbb{p}$	$\mathbb{CL}^{\neg}, \mathbb{M}_2^{\neg}$	Y
$\mathsf{IL}^{\to} \bullet \top\!\!\!\bot$	**(mp) (a1) (a2)**	**(wxc) (di) (wde)**	$**$	$\mathbb{G}_3^{\neg}, \mathbb{M}_2^{\neg}$	Y
$\mathsf{MP}_1 \bullet \top\!\!\!\bot$	**(mp) (a1)**	**(wxc)**	$\mathbb{MP}_1 \star \mathbb{B}\mathbb{o}\mathbb{t}\mathbb{o}\mathbb{p}$	$\mathbb{MP}_1, \mathbb{M}_2^{\neg}$	Y
$\mathsf{MP} \bullet \top\!\!\!\bot$	**(mp)**	none	$\mathbb{MP}_{\circ} \star \mathbb{B}\mathbb{o}\mathbb{t}\mathbb{o}\mathbb{p}$	$\mathbb{F}^{\neg}$	Y
$\mathsf{MP}_{\circ} \bullet \bot$	**(cmp) (xf)**	**(gxc)**	$\mathbb{MP}_{\circ} \star \mathbb{B}\mathbb{o}\mathbb{t}$	$\mathbb{F}^{\neg}$	Y
$\mathsf{MP}_{\circ} \bullet \top\!\!\!\bot$	**(cmp)**	none	$\mathbb{MP}_{\circ} \star \mathbb{B}\mathbb{o}\mathbb{t}\mathbb{o}\mathbb{p}$	$\mathbb{F}^{\neg}$	Y

Fig. 4 Logics and results ($*$ means that the logic is characterizable by a single infinite matrix, but not by a finite set of finite Nmatrices, and $**$ means that the logic is characterizable by a single infinite Nmatrix but not by a finite set of finite Nmatrices.)

5 Concluding Remarks

The main purpose of this paper has been to illustrate the usefulness of some recently-developed techniques in the study of logics defined via Hilbert-style rules. We have focused on the negation fragments of logics which result from different possible choices of well-known rules involving the connectives $\{\to, \bot\}$, with a few variations. Our results are summarized in Fig. 4. Further, all the negations defined by making $\neg p := p \to \bot$, in these logics are characterized by subsets of the rules in Fig. 3 below.

We have shown that all the negation-only logics obtained are characterized by finite (N)matrices, which immediately implies that their decision problem is in **coNP**. As a matter of fact, it is not hard to see that all of these logics are actually in **P**.

We have quite different complexities for the environment logics from which we have obtained the negations. It is easy to see that MP, $\mathsf{MP}_{\circ}$, $\bot$ and $\top\!\!\!\bot$ are in **P**. Since **P** is closed under disjoint fibring [4], it follows that the logics $\mathsf{MP} \bullet \bot$, $\mathsf{MP}_{\circ} \bullet \bot$, $\mathsf{MP} \bullet \top\!\!\!\bot$ and $\mathsf{MP}_{\circ} \bullet \top\!\!\!\bot$ are also in **P**. Notice that, while $\mathsf{IL}^{\to} \bullet \bot$ and IL are in **PSPACE**, and $\mathsf{CL}^{\to} \bullet \bot$ is in **coNP**, in $[\mathsf{IL}^{\to} \bullet \bot, \mathsf{CL}^{\to} \bullet \bot] \cup [\mathsf{IL}^{\to} \bullet \bot, \mathsf{IL}]$ there are even undecidable logics.

While the topics we have touched on certainly deserve to be pursued in a more systematic fashion, we hope that our playful exploration may be useful and insightful on the nature of negations and their paraconsistent character.

The game we played is obviously not limited to the connectives and logics considered here, and can be recast in more general contexts. One such obvious possibility is the study of modalities, which have well-known connections with negations [7]. It is however fair to mention that we are still far from completely understanding the mechanics of Hilbert-style calculi. In [12] we describe the semantics for disjoint fibring, but we still fall short of the general case.

Acknowledgements The first two authors thank the support of the FEDER/FCT project PEst-OE/EEI/LA0008/2013 of Instituto de Telecomunicações. The first author also acknowledges the FCT postdoctoral grant SFRH/BPD/76513/2011. Finally, we would like to thank Fey Liang for pointing out a deadly mistake in an earlier version of the paper.

References

1. Abbott, J.C. 1967. Implicational algebras. *Bulletin Mathématique de la Société des Sciences Mathématiques de la République Socialiste de Roumanie* 11: 3–23.
2. Avron, A., and I. Lev. 2005. Non-deterministic multiple-valued structures. *Journal of Logic and Computation* 15: 241–261.
3. Avron, A., and A. Zamansky. 2011. Non-deterministic semantics for logical systems. In *Handbook of philosophical logic*, vol. 16, 2nd ed, ed. D.M. Gabbay, and F. Guenthner, 227–304. Heidelberg: Springer.
4. Caleiro, C., and S. Marcelino. 2016. Decidability and complexity of fibred logics without shared connectives. *Logic Journal of IGPL* 24: 673–707.
5. Caleiro, C., S. Marcelino, and U. Rivieccio. 2018. Characterizing finite-valuedness. *Fuzzy Sets and Systems* 345: 113–125.
6. Carnielli, W., and J. Marcos. 2002. A taxonomy of c-systems. In *The logical way to the inconsistent*, ed. M. Coniglio, W. Carnielli, and I.M. D'Ottaviano, 1–94. Baco Raton: CRC Press.
7. Dodó, A., and J. Marcos. 2014. Negative modalities, consistency and determinedness. *ENTCS* 300: 21–45.
8. Galatos, N., P. Jipsen, T. Kowalski, and H. Ono. 2007. *Residuated lattices: an algebraic glimpse at substructural logics*, vol. 151., Studies in logic and the foundations of mathematics Amsterdam: Elsevier.
9. Gottwald, S. 2001. *A treatise on many-valued logics*, vol. 9., Studies in logic and computation Baldock: Research Studies Press.
10. Hájek, P. 1998. *Metamathematics of fuzzy logic*, vol. 4., Trends in logic-studia logica library Dordrecht: Kluwer Academic Publishers.
11. *International Journal of Mathematics and Mathematical Sciences*2002. A Note on Homomorphisms of Hilbert Algebras. 29: 55–61. https://doi.org/10.1155/S0161171202011134.
12. Marcelino, S., and C. Caleiro. 2016. On the characterization of fibred logics, with applications to conservativity and finite-valuedness. *Journal of Logic and Computation*. https://doi.org/10.1093/logcom/exw023.
13. Marcelino, S., and C. Caleiro. 2017. Disjoint fibring of non-deterministic matrices. In *Logic, language, information, and computation*, ed. J. Kennedy, and R. de Queiroz, 242–255. Heidelberg: Springer.
14. Marcos, J. 2009. What is a non-truth-functional logic? *Studia Logica* 92: 215–240.
15. Rasiowa, H. 1974. *An algebraic approach to non-classical logics*., Studies in logic and the foundations of mathematics Amsterdam: North-Holland.
16. Statman, R. 1979. Intuitionistic propositional logic is polynomial-space complete. *Theoretical Computer Science* 9: 67–72. https://doi.org/10.1016/0304-3975(79)90006-9.
17. Wójcicki, R. 1988. *Theory of logical calculi. basic theory of consequence operations*, vol. 199., Synthese library Dordrecht: Reidel.

Correction to: Contradictions, from Consistency to Inconsistency

Walter Carnielli and Jacek Malinowski

Correction to:
W. Carnielli and J. Malinowski (eds.), *Contradictions, from Consistency to Inconsistency*, Trends in Logic 47, https://doi.org/10.1007/978-3-319-98797-2

The book was inadvertently published with chapter author's incorrect family name. This information has been updated from "First Name: María and Family Name: del Rosario Martínez-Ordaz" to "First Name: Maria del Rosario and Family Name: Martínez-Ordaz" in the initially published version of Chapters "The Possibility and Fruitfulness of a Debate on the Principle of Non-contradiction" and "Keeping Globally Inconsistent Scientific Theories Locally Consistent". The correction chapters and book have been updated with the changes.

The updated version of these chapters can be found at
https://doi.org/10.1007/978-3-319-98797-2_3
https://doi.org/10.1007/978-3-319-98797-2_4

W. Carnielli and J. Malinowski (eds.), *Contradictions, from Consistency to Inconsistency*, Trends in Logic 47, https://doi.org/10.1007/978-3-319-98797-2_15