Truth, Paradoxes, and Partiality

A Study on Semantic Theories of Naïve Truth

Lorenzo Rossi

New College
University of Oxford

A thesis presented for the degree of
Doctor of Philosophy
Michaelmas 2015

A Benedetta, ai miei genitori e a mia sorella

Abstract

Truth, Paradoxes, and Partiality
Lorenzo Rossi, New College, A thesis presented for the degree of Doctor of Philosophy, Michaelmas 2015

This work is an investigation into the notion of *truth*. More specifically, this thesis deals with how to account for the main features of truth, with the interaction between truth and fundamental linguistic elements such as connectives and quantifiers, and with the analysis and the solution of truth-theoretic paradoxes.

In the introductory Chapter 1, I describe and justify the approach to truth I adopt here, giving some general coordinates to contextualize my work.

In Part I, I examine some theories of truth that fall under the chosen approach. In Chapter 2, I discuss a famous theory of truth developed by Saul Kripke. Some difficulties of Kripke's theory led several authors, notably Hartry Field, to emphasize the importance of a *well-behaved conditional connective* in conjunction with a Kripkean treatment of truth. I articulate this idea in a research agenda, which I call *Field's program*, giving some conditions for its realizability. In Chapter 3, I analyze the main theory of truth proposed by Field to equip Kripke's theory with a well-behaved conditional, and I give a novel analysis of its shortcomings. Field's theory is remarkably successful but is technically and intuitively very complex, and it is unclear whether Field's conditional is a plausible candidate for a philosophically useful conditional. Moreover, Field's treatment of "determinate truth" and his handling of many kinds of paradoxes is not fully satisfactory.

In Part II, I develop some new theories that capture the main aspects of the notion of truth and, at the same time, give a philosophically interesting meaning to connectives and quantifiers – in particular, they yield a strong and conceptually significant conditional. The theory proposed in Chapter 4 extends the inductive methods employed in Kripke's theory, showing how to adapt them to non-monotonic connectives as well. There, I also develop and defend a new, theoretically fruitful notion of gappiness. The theory proposed in Chapter 5 (and discussed further in Chapter 6), instead, employs some graph-theoretic intuitions and tools to provide a new model-theoretic construction. The resulting theory, I argue, provides a nice framework to account for the interaction between truth, connectives, and quantifiers, and it is flexible enough to be applicable to several interpretations of the logical vocabulary. Some new technical results are established with this theory as well, concerning the interplay between every Łukasiewicz semantics and some interpretations of the truth predicate, and concerning the handling of determinate truth. Finally, the theory developed in Chapter 5 provides articulate and telling solutions to truth-theoretical paradoxes.

Acknowledgements

Until recently, I held a false belief about acknowledgements, namely that they are easy to write. When I started to write this section, however, it soon became clear to me that the quality and the quantity of support that I received during my doctorate, on an academic and a personal level, requires far more recognition than I am able to express. Nevertheless, despite my inability to thank all those whom I want to thank, and to do so appropriately, I deeply appreciated the opportunity to think of all the attention I received during my doctorate, to reflect on how fortunate I have been, and to try to convey my gratitude at the best of my possibilities.

I owe to Volker Halbach, my supervisor, an amount of thanks that is not easily expressed. Volker has been an inspiring presence during my doctorate, with his continuous advices, encouragement, and support. If I look back to how I was when I began my doctorate, I am struck by how much Volker helped me to develop a sounder methodological approach, to look for better tools and perspectives to deal with the problems I encountered, to inquire more deeply into the implications of a philosophical claim or of a logical demonstration. Volker taught me a lot on theories of truth and other topics, but what I will remember most of his work as my supervisor will be the feeling of being introduced into active research – not passively "shown" a philosopher and logician at work – but the awareness of being given the best conditions to work in an independent way. Volker was able to provide such conditions through a tireless and articulated work, and I am not sure I understand fully how difficult it is to supervise a student as thoroughly as he does. I hope that my efforts, most of which are in this thesis, could show him how deeply I appreciated his guidance and the opportunity of studying with him.

I would also like to thank the former members of Volker's research group *Inexpressibility and Reflection in the Formal Sciences*, namely Graham Leigh, James Studd, Kentaro Fujimoto, and especially Carlo Nicolai. They made me feel at home in the Faculty and at ease with my research. I asked their advice in far more occasions than I can remember, and they were always ready to talk about non-classical logic and weird models with me. In particular, I am profoundly grateful to Carlo for his support, his closeness, and his friendship.

My secondary supervisor, Tim Williamson, has also been extraordinarily helpful over the course of my doctoral work. He has always been very supportive and happy to read my drafts, sharing with me his insightful comments. In our discussions, he illuminated several aspects of my work that had potential for developments, and he was fundamental in enabling me to grasp the importance of relating specific results and philosophical points to the overall aims of my work.

I wish to express my gratitude also to all the professors, researchers, and students at the Faculty of Philosophy of the University of Oxford with whom I could discuss about philosophy and logic, and from whom I could learn so much: in particular, I would like to thank James Studd, Jeff Ketland, Dan Isaacson, Alex Paseau, Paolo Fait, Andreas Ditter, Beau Madison Mount, Andy Yu, Harvey Lederman, Peter Fritz, David Schroeren, Samuel Meister, and Neil Dewar. Lorenzo

Greco deserves a host of thanks for making my life as a graduate student much easier and much more enjoyable. The Faculty of Philosophy at the University of Oxford is a wonderful environment to carry on doctoral work: I wish to thank all those who make it such a great working place.

During my time in Oxford, I was lucky enough to present my work to and discuss it with Leon Horsten, Philip Welch, Hannes Leitgeb, and Ole Hjortland: to all of them, my sincerest thanks. A special thanks goes to Marianna Antonutti Marfori and Benedict Eastaugh for the lovely time spent together and for hosting me during a visit to Munich.

During the Fall 2014, I spent a semester as a visiting PhD student at New York University. I am immensely grateful to Hartry Field for allowing my visit and for supervising my research while I was at NYU. Hartry has always been incredibly kind and encouraging, discussing my work with me on numerous occasions. He has always been full of thought-provoking suggestions, and he raised several issues that helped me to improve my work in a way that can hardly be overestimated. I am also grateful to several members of the NYU Philosophy Department for their precious comments, and in particular to Tim Maudlin, Paul Horwich, Kit Fine, Andreas Ditter (now as an NYU student), Andy Yu (as a visiting NYU student), Daniel Hoek, Andrés Soria Ruiz, and Louis Rouillé. Special thanks go to Harvey Lederman (now as an assistant professor and post-doc researcher at NYU) for his invaluable feedback.

During my time in New York, I had the fortune to discuss my research with several other professors and students at other Universities: I would like to thank Graham Priest, Melvin Fitting, Achille Varzi, Haim Gaifman, and Bruno Whittle. I am also grateful to Emi and Hitoshi Omori for their kindness and friendship.

Huge thanks are due to the professors who taught me as an undergraduate and as a master's student at the University of Florence, and especially to Andrea Cantini, Pierluigi Minari, Elena Castellani, and Gianluca Garelli. In particular, I wish to express all my gratitude to Andrea Cantini: not only he was an extraordinary supervisor for my master's thesis, but he has also been constantly in contact with me during my doctorate, and he has always been happy to discuss my work every single time I came back to Italy, offering his invaluable feedback and advices. To Pietro, Daniele, Rossella, Paolo, Tommaso, and my other former classmates in Florence, my thanks for their continuing presence and support.

I am deeply indebted to Charlotte Werndl for supporting my application for a post-doc at the Department of Philosophy at the University of Salzburg and for making me feel welcome and at home there; I would to thank also all the other members of the Department with whom I have had the pleasure to talk about my research, including Julien Murzi, Christopher Gauker, and Johannes Brandl.

My doctoral studies would not have been possible without the generous support of several scholarships and grants: I am very grateful for the Arts and Humanities Research Council (AHRC) Studentship, the Scatcherd European Scholarship, the AHRC Research Training Support Grant for Overseas Study Visit, and several travel grants from the Faculty of Philosophy and from New College.

New College proved to be a great environment to work, and made it easy to meet very special people who made my days in Oxford very happy: amongst them, I would like to mention Jim, John, Julian, Andreas, Arnold, Kira, Ricky, Amin, Mattia, Viviana, Bruno, Giada, Gianluca, and Mira. Mattia deserves special thanks for at least two reasons: first, because he endured two hours of explanation on what are theories of truth and he never retaliated with a similarly detailed explanation of whatever part of Astrophysics he works in; second, because he is almost as bad at playing tennis as I am, thus being an ideal sport partner.

My profoundest gratitude goes to Viviana, to Keri and Ben, and to Nicole, Iginio, and little Carolina, for having supported Benedetta in a period that could have been very difficult for her and that turned out to be happy and productive instead: that gave me a lot of serenity for my work.

My old bunch of friends from back home have been a constant presence of all these years, and I would like to thank Giacomo, Lorenzo, Giulio, Luca, and of course Elisa and Lorenzo for their closeness and affection.

I owe an immense debt of gratitude to Michele: I would like to thank him for being always next to me, ready to listen to me and to give me all his attention and encouragement. He showed me that some friendships are so deep-rooted that distance or difficulties can do nothing but strengthen them.

I wish to thank my family for their inexhaustible support, in particular my parents, my sister and her partner, as well as my numerous relatives and Benedetta's family: they all made all our returns to Italy into delightful periods.

Finally, Benedetta provided the most invaluable support I could hope for, becoming my wife. Since every attempt to thank her would inevitably fail to convey my gratitude, I can only express the hope to be as caring for her as she is for me.

Oxford
October 30, 2015

I wish to thank my examiners, Jeff Ketland and Leon Horsten, for their careful review of my work, for their precious feedback and suggestions, and for making my *viva* as useful and enjoyable as I could hope for.

Salzburg
April 7, 2016

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Aim of the thesis

This thesis is a contribution to the study of the notion of *truth*. Understanding the notion of truth has a remarkable philosophical significance, since truth plays a central role in philosophical investigations (e.g. in philosophy of language, epistemology, and more), in scientific theories, and in the ordinary linguistic practice.

Few, if any, notions have been investigated as extensively and deeply throughout the history of philosophy as truth, so any non-encyclopedic work on this topic has to carve out a relatively small area of investigation and concentrate on it. Fortunately, recent developments in philosophy and logic have delivered us a well-established field of research, known as *theories of truth*, that makes it easier to design specific enquiries, within the countless possibilities offered by the notion of truth. Theories of truth have multiple objectives but, somewhat oversimplifying, they are associated by the aims of exploring the notion of truth, of modeling the behavior of the expression "… is true" (in formal or natural languages), and of developing solutions to truth-theoretic paradoxes.

Despite the specialization achieved by contemporary theories of truth, the field remains very large, with ramifications that touch linguistics, mathematical logic, computer science, and still more disciplines. Moreover, even within the philosophically oriented investigations on truth, there are plenty of approaches to choose from. Charting all such approaches would be beyond the scope of this introduction, and it would not be functional to present the framework I adopt and the nature of my contribution. So, I will contextualize my work via some conceptual coordinates, corresponding to some major options in thinking about truth.

Perhaps the first point where theories of truth diverge is, unsurprisingly, the interpretation of the truth predicate. An important tradition maintains that the fundamental trait of the notion of truth is captured by a view along the following lines.

(NAÏVETÉ) For some notion of equivalence, for every sentence A of the language under consideration (that includes the truth predicate), A and "«A» is true" are equivalent.[1]

The formulation of (NAÏVETÉ) is not fully precise but is evidently intelligible. I will count as "naïve" all the views that adopt a version of (NAÏVETÉ), which seems general enough to span every theory customarily considered naïve.

The idea that truth is characterized by (NAÏVETÉ) is chiefly supported by the use of the truth predicate, from ordinary languages to the scientific discourse. It seems hard to deny that some version of (NAÏVETÉ) is at the heart of truth, if we consider the role and the behavior of this notion across a variety of contexts. So, as a first coordinate, I will say that my work is in the area of *theories of naïve truth*.[2]

[1] This tradition lives mainly within the truth-theoretic *deflationism*. I cannot introduce this position here: major works on it include [Field 1994a], [Field 1994b], [Horwich 1998], [Shapiro 1998], [Ketland 1999], [Field 1999], [Halbach 1999], [Halbach 2001a], [Halbach 2001b], [Horsten 2009].

[2] Finding convincing justifications for the naïve perspective on truth, or deciding whether truth possesses other theoretically relevant feature beyond naïveté is not amongst the objectives of this work. The literature on truth-theoretic deflationism is the primary source for this discussion, and I refer the reader to that. I take it that (NAÏVETÉ) is intuitive enough to avoid the need of advocacy: the problems surrounding (NAÏVETÉ), as we shall see, are more concerned with the ways to account for it.

1.2 Paradoxes

Naïve truth is not an easy position to embrace, as several *paradoxes* or *antinomies* arise from its acceptance – I will use these terms as synonymous.[3] The more time-honored truth-theoretic paradox is undoubtedly the *liar paradox*, that can be presented informally as follows. Consider the following sentence.

(L) The sentence labelled with "(L)" on page 3 of this thesis is not true.

Consider now the following "liar reasoning".[4] Suppose that (L) is true. By (NAÏVETÉ), what (L) says is the case, so the sentence labelled with "(L)" on page 3 of this thesis is not true; but such sentence is (L) itself, so (L) is not true. Suppose that (L) is not true. But then what (L) says is the case and so, by (NAÏVETÉ), (L) is true. Since (L) is either true or not true, by the above reasoning it is true and not true, which is a contradiction, and from contradictions everything follows.

The liar reasoning, one might think, shows that if one accepts (NAÏVETÉ), she accepts everything, so (NAÏVETÉ) has to be rejected. Several authors, however, advocated the need to resist this conclusion, looking for a way to rescue (NAÏVETÉ) from triviality – this effort is motivated by the compelling nature of (NAÏVETÉ). After all, several steps can be questioned in the liar reasoning: is it the case that (L) is either true or not true? is it the case that from contradictions everything follows?

These questions suggest that the analysis of naïve truth can be carried out on the ground of the principles we accept in reasoning and of the interpretation we give to linguistic elements such as negation or disjunction, i.e. on *logical* grounds. We met another divide within theories of truth, which can be developed *formally* or *informally*. Formal theories employ logico-mathematical techniques. As a second coordinate, the present work belongs to the area of *formal theories of naïve truth*.

[3] For an introductory discussion of paradoxes and some history, see [Cantini 2009].
[4] I am roughly following [McGee 1991]'s presentation here.

Several reasons suggest a formal approach. A logico-mathematical approach can provide us precise instruments to analyze sentences and to control the principles we adopt, in order to find solutions to paradoxes. As the liar shows, moreover, only simple fragments of English are needed to formulate some paradoxes, and some mathematical languages can be used to reconstruct their functioning: so, it is reasonable to study paradoxes in a minimally simple setting, disregarding the irrelevant intricacies of natural languages. Nevertheless, the formal languages employed in theories of truth can be generally translated into natural languages, so the results of formal investigations may be valuable for natural languages as well.

Formal theories of truth can be roughly divided in two interconnected kinds: *axiomatic* theories of truth study principles that the notion of truth should respect; *semantic* theories of truth explore models of languages featuring the truth predicate.[5] As a third coordinate, this work deals with *semantic theories of naïve truth*.

I recall now some technical notions and notational conventions. For reasons of space, I cannot introduce them from scratch: since this thesis is a research work and not a handbook, much of it has to be devoted to give original contributions to semantic theories of naïve truth. So, I will have to assume familiarity with some logico-mathematical notions, but I will always provide references for them.

1.3 Languages and notational conventions

Formal investigations on truth require a formal language where the truth predicate is modeled. Following a growing body of research, I will study the language of first-order arithmetic, call it $\mathscr{L}$, expanded with a unary predicate T for "... is true": indicate this language with $\mathscr{L}_T$. I will also consider other languages, obtained supplementing $\mathscr{L}_T$ with new connectives or quantifiers. I will sometimes refer to

[5]For an overview of formal approaches to truth, see [Sheard 1994]. A classic introduction to semantic theories of truth is [Visser 2002], and a comprehensive text of model theory is [Hodges 1993]. For an encompassing study of axiomatic theories of truth, see [Halbach 2011].

Peano Arithmetic, formulated in $\mathscr{L}_T$.[6] The language of arithmetic is chosen mainly because it is simple and it has been extensively studied, and because (NAÏVETÉ) and several paradoxes can be modeled in it in precise ways.

The primitive connectives and quantifiers of $\mathscr{L}_T$ are $\neg$, $\wedge$, $\vee$ and $\forall$. Terms and formulae of $\mathscr{L}_T$ are defined as usual. Formulae with no free variables are called "sentences", whereas "term" refers to closed and open terms alike. I use $s_1, t_1, \ldots, s_n, t_n, \ldots$ to range over $\mathscr{L}_T$-terms. Lowercase Greek letters late in the alphabet (e.g. φ, ψ, χ but not ω) refer to $\mathscr{L}_T$-formulae; uppercase Greek letters early in the alphabet (Γ, Δ) refer to sets of $\mathscr{L}_T$-formulae. $\varphi \supset \psi$ abbreviates $\neg\varphi \vee \psi$, $\varphi \equiv \psi$ abbreviates $(\varphi \supset \psi) \wedge (\psi \supset \varphi)$, and $\exists x \varphi(x)$ abbreviates $\neg\forall x \neg\varphi(x)$.

Fix a coding system for $\mathscr{L}_T$.[7] φ in Gödel quotes, $\ulcorner\varphi\urcorner$, denotes the numeral of the code of φ. If f is a (primitive recursive) function, $\underset{\cdot}{f}$ denotes f as represented in $\mathscr{L}_T$.[8] Expressions of $\mathscr{L}_T$ are identified with their codes. $TER_{\mathscr{L}_T}$, $CTER_{\mathscr{L}_T}$, $SENT_{\mathscr{L}_T}$, $FOR_{\mathscr{L}_T}$ indicate the sets of (codes of) terms, closed terms, sentences and formulae of $\mathscr{L}_T$, respectively. "$\varphi \in \mathscr{L}_T$" stands for "$\varphi$ is a $\mathscr{L}_T$-sentence".

Some letters are reserved for sentences that formalize famous paradoxes in $\mathscr{L}_T$:

1. λ is called the *liar sentence* and is the sentence $\neg Tt_\lambda$, where t_λ denotes the code of λ, i.e. $t_\lambda = \ulcorner\lambda\urcorner = \ulcorner\neg Tt_\lambda\urcorner$.[9]

2. $\kappa_\supset$, *Curry's material sentence*, is the sentence $Tt_{\kappa_\supset} \supset 0 = 1$ such that $t_{\kappa_\supset} = \ulcorner\kappa_\supset\urcorner = \ulcorner Tt_{\kappa_\supset} \supset 0 = 1\urcorner$.

3. τ, the *truth-teller*, is the sentence Tt_τ such that $t_\tau = \ulcorner\tau\urcorner = \ulcorner Tt_\tau\urcorner$.

4. μ, *McGee's sentence*, is the sentence $\neg\forall n T\ulcorner \underset{\cdot}{g}(n, t_\mu)\urcorner$ such that $t_\mu = \ulcorner\mu\urcorner = \ulcorner\neg\forall n T\ulcorner \underset{\cdot}{g}(n, t_\mu)\urcorner\urcorner$. g is the function that takes as input a natural number n

[6] Most of the notational conventions I adopt are from [Halbach 2011]. For $\mathscr{L}$ and Peano Arithmetic, see [Kaye 1991] or [Hájek and Pudlák 1998].

[7] E.g. of the kind described in [van Dalen 2012], Chapter 8.1 (details are unimportant).

[8] For arithmetical representation and quantification over variables occurring in sentences within Gödel quotes, see [Feferman 1991] §2.1, [Halbach 2011] Chapter 5, [van Dalen 2012], Chapter 8.5.

[9] This characterization is from [Halbach 2011], p. 160.

and (the code of) a sentence $\varphi \in \mathscr{L}_T$, and yields as result the sentence made of n iterations of T in front of φ.[10]

Lowercase Greek letters early in the alphabet $(\alpha, \beta, \gamma, \delta)$ are for ordinals, with the exceptions of ω (the least infinite ordinal, for convenience identified with the set of non-negative integers) and ω_1^{CK} (the least non-recursive ordinal). Ordinals are underlined (e.g. $\underline{0}$, $\underline{\alpha}$) if represented in $\mathscr{L}_T$ and overlined (e.g. $\overline{0}$) to disambiguate them from $\mathscr{L}_T$-terms. The class of ordinals is indicated with Ord.[11]

Lowercase Greek letters middle in the alphabet (ζ, η, ϑ) are for formulae of $\mathscr{L}_T$ (or superlanguages of it) with set variables allowed (i.e. the language of the theory of inductive definitions for arithmetic). Sets of natural numbers are indicated with P, Q, R, set variables are indicated with $S_1, \dots, S_i, \dots$, and operators on them are indicated with uppercase Greek letters late in the alphabet (Φ, Ψ, Υ).[12]

"s.t." abbreviates "such that", "iff" abbreviates "if and only if", "IH" abbreviates "inductive hypothesis", $\Box$ marks the end of a proof, and $\blacktriangle$ marks the end of a definition, a characterization, a remark, or an example. If a result is stated without proof, it is a well-established result, or it is immediate from what precedes it.

1.4 Varieties of naïveté, non-classical semantics

Adopting a formal approach to truth, we can give precise formulations of (NAÏVETÉ) and of the equivalence it postulates. I am interested in semantic formulations; in particular, I will deal with three realizations of (NAÏVETÉ), often debated in the literature. Rigorous formulations of these theses are theory-dependent, but we can give semi-formal templates that are applicable to several theories.[13] Suppose that

[10] See [McGee 1985]; I followed [Halbach 2011]'s formalization.

[11] For ordinal numbers, see [Jech 2002], Chapter 2; for ω_1^{CK}, see [Rogers 1987], Chapter 11.7.

[12] For the theory of inductive definitions, see [Moschovakis 1974] and [Aczel 1977].

[13] These characterizations are in order of increasing strength: usually, theories that validate versions of naïveté down in the list also validate the versions up in the list, but not *vice versa*.

a semantic theory of truth yields an evaluation function $\mathscr{E}$, i.e. a function from the sentences of $\mathscr{L}_T$ (or superlanguages of it) to a value space.

Characterization 1.4.1. (Weak naïveté)

$\mathscr{E}$ respects *weak naïveté* if, for every $\varphi \in \mathscr{L}_T$:

$$\mathscr{E}(\varphi) =^{\star} \mathscr{E}(T\ulcorner\varphi\urcorner),$$

where $=^{\star}$ is *some* notion of equality (not necessarily identity).[14] ▲

Characterization 1.4.2. (Intersubstitutivity of truth (Field))[15]

$\mathscr{E}$ respects *intersubstitutivity of truth* if for every $\varphi, \psi, \chi \in \mathscr{L}_T$, if ψ and χ are alike except that ψ has an occurrence of φ of $\mathscr{L}_T$ where χ has an occurrence of $T\ulcorner\varphi\urcorner$:

$$\mathscr{E}(\psi) =^{\star} \mathscr{E}(\chi),$$

where $=^{\star}$ is *some* notion of equality, as indicated above. ▲

Characterization 1.4.3. (Strong naïveté)

$\mathscr{E}$ respects *strong naïveté* if, for every $\varphi \in \mathscr{L}_T$ (or superlanguages of it):

$$\mathscr{E}(\varphi \equiv^{\star} T\ulcorner\varphi\urcorner) = \mathbf{d},$$

where $\mathbf{d}$ is the/a designated value and $\equiv^{\star}$ is *some* biconditional.[16] So, $\mathscr{E}$ respects *strong naïveté* if it validates the *Tarski biconditionals* for some biconditional. ▲

$=^{\star}$ and $\equiv^{\star}$ suggest a departure from classical, total evaluations for the sentences of $\mathscr{L}_T$ (or superlanguages of it). In fact, one of the most celebrated results

[14] In practice, however, only the usual notion of identity and Kleene equality will be used.

[15] Slight modification of [Field 2008], p. 12.

[16] This formulation does not says exactly *what is a biconditional*: this issue, it seems to me, should not be addressed in general, but only in the context of specific theories, so the characterization remains intelligible. For a quick discussion on this, see [Field 2008], p. 276.

on formal truth, *Tarski's Theorem on the Undefinability of Truth*, shows that there is no *classical and total* evaluation function $\mathscr{E}$ for the sentences of $\mathscr{L}_T$ (or superlanguages of it) that respects weak naïveté, intersubstitutivity of truth, or strong naïveté – to see this, let $\mathscr{E}$ be classical and total, and consider $\mathscr{E}(\lambda)$.[17]

Devising semantic theories of naïve truth consists largely in finding evaluations that respect a form of naïveté while providing a satisfactory assignment of semantic values to the sentences of $\mathscr{L}_T$ (or superlanguages of it). There is little disagreement on how to assign values to arithmetical sentences (we have the standard model of arithmetic for this): the real challenge is to keep (NAÏVETÉ) while giving a nice account of connectives and quantifiers and a nice semantic value assignment to sentences, and providing telling analyses of and satisfactory solutions to truth-theoretic paradoxes. This is also the purpose of the theories I develop here.

Here is an evaluation scheme that appears often in theories of truth.

Definition 1.4.4. (Strong Kleene evaluation scheme)[18]

The *strong Kleene evaluation scheme* for $\mathscr{L}_T$ (but it is similar for superlanguages of it) is described by the functions $\mathscr{E}_{\mathrm{K3}} : SENT_{\mathscr{L}_T} \longmapsto \{0, 1/2, 1\}$ that respect the following conditions: for every $\varphi, \psi \in \mathscr{L}_T$, and every $\chi(x) \in FOR_{\mathscr{L}_T}$,

$$\begin{aligned}
\mathscr{E}_{\mathrm{K3}}(\neg\varphi) &= 1 - \mathscr{E}_{\mathrm{K3}}(\varphi), \\
\mathscr{E}_{\mathrm{K3}}(\varphi \wedge \psi) &= \min(\mathscr{E}_{\mathrm{K3}}(\varphi), \mathscr{E}_{\mathrm{K3}}(\psi)), \\
\mathscr{E}_{\mathrm{K3}}(\varphi \vee \psi) &= \max(\mathscr{E}_{\mathrm{K3}}(\varphi), \mathscr{E}_{\mathrm{K3}}(\psi)), \\
\mathscr{E}_{\mathrm{K3}}(\forall x \chi(x)) &= \inf\{\mathscr{E}_{\mathrm{K3}}(\chi(t_n)) : \text{ for all } t_n \in TER_{\mathscr{L}_T}\}.
\end{aligned}$$

▲

[17] By *classical and total*, I mean that $\mathscr{E}$ respects the classical Tarskian evaluation clauses and that, for every sentence of $\mathscr{L}_T$ (or superlanguages of it), $\mathscr{E}$ gives exactly one classical semantic value to it. The *locus classicus* for Tarski's result is [Tarski 1935]. Tarski's Theorem is usually presented for some formal system, but of course I am interested in its semantic consequences.

[18] For strong Kleene semantics, see [Malinowski 1993], [Urquhart 2002], [Blamey 2002].

Strong Kleene semantics is employed often to address truth-theoretic paradoxes. In the liar reasoning above, we saw that that both accepting that the liar sentence is true and accepting that it is not true lead to problems. So, a natural option is refusing to do any of these two moves. In strong Kleene semantics the law of excluded middle is not validated: the instances of $\varphi \vee \neg\varphi$ where φ is assigned value $1/2$ receive value $1/2$ as well (1 is the only designated value). Moreover, if we give λ value $1/2$, this saves us from the embarrassing position of accepting $\lambda \vee \neg\lambda$: as Horsten points out,[19] everyone accepting $\lambda \vee \neg\lambda$ accepts $(\lambda \wedge \neg T\ulcorner\lambda\urcorner) \vee (\neg\lambda \wedge T\ulcorner\lambda\urcorner)$ – unless she accepts a sentence φ without accepting $\varphi \wedge \varphi$ or $\neg\neg\varphi$. But $(\lambda \wedge \neg T\ulcorner\lambda\urcorner) \vee (\neg\lambda \wedge T\ulcorner\lambda\urcorner)$ is obviously undesirable, as it shows that we accept something untrue or that we deny something true.

Rejecting excluded middle is not only a promising way to avoid paradoxes and related problems, but it seems also to offer interesting perspectives for their studies: if we are not forced to classify everything as true or not, we can enquire into the peculiar status of paradoxical sentences, looking for novel and satisfactory ways to account for them. For this and other reasons, approaches that reject the law of excluded middle have been extensively investigated, under the name of *paracomplete* approaches. I can now give the last coordinate to introduce this thesis, stating that it is a study of and a contribution to the field of *paracomplete semantic theories of naïve truth*. In the first part of this work, I analyze some major paracomplete semantic theories of naïve truth, and in the second part I provide new paracomplete semantic theories of naïve truth. The main objective of the theories I propose is to develop a satisfactory understanding of naïve truth, in its interaction with connectives and quantifiers, an acceptable assignment of semantic values to sentences of $\mathscr{L}_T$ (or superlanguages of it), and an appropriate treatment of truth-theoretic paradoxes, from the perspective I introduced in this section.

[19] See [Horsten 2011], pp. 127-128.

Part I

Analysis of Some Semantic Theories of Naïve Truth

Chapter 2

Kripke's Theory

2.1 Introduction

The celebrated paper "Outline of a Theory of Truth" published by Saul Kripke ([1975]) is considered by many authors to be the single greatest contribution to theories of truth since Tarski's groundbreaking essay [Tarski 1935]. In this chapter, I will present and discuss Kripke's theory, relating it to the aims of my work.

From a historical point of view, it should be noted that constructions essentially similar to some of Kripke's models had already been developed by Paul C. Gilmore for partial set theory,[1] and by Robert L. Martin and Peter Woodruff for naïve truth,[2] even if the version of Martin and Woodruff is actually just a special case of Kripke's construction. Kripke refers to Martin and Woodruff's work, relating their results to his ones, and briefly hints at approaches similar to his own in the area of set-theoretic paradoxes, mentioning Gilmore, Fitch, and Feferman.

However, neither the question of who invented Kripke's kind of models, nor the mathematical aspects of such models are central in assessing the importance of Kripke's work. From the mathematical standpoint, in fact, the constructions

[1] See [Gilmore 1967] and [Gilmore 1974].

[2] See [R.L. Martin and Woodruff 1975].

employed by Gilmore, Martin and Woodruff, or Kripke himself are all relatively straightforward applications of some facts of the theory of inductive definitions.[3] Moreover, unlike the work of its predecessors, Kripke's paper remains fundamental for at least three reasons: (i) its convincing critique of the approaches to truth-theoretic paradoxes that postulate a Tarski-hierarchy of truth predicates;[4] (ii) the general study of a kind of semantic construction across various evaluation schemes;[5] (iii) the classification of the various kinds of models (fixed points) obtained in such constructions, and their philosophical applications.

Let me now turn to an informal description of Kripke's theory.

2.2 Kripke's theory: informal presentation

Consider a language such as $\mathscr{L}_T$: we have a good idea of how to interpret its arithmetical part, by looking at the standard model of arithmetic. $\mathbb{N}$ gives a denotation to all the closed $\mathscr{L}$-terms and, most importantly for my concerns, a semantic value to every $\mathscr{L}$-sentence. This, however, does not put us in the position to interpret the truth predicate and the sentences containing it. In order to do that, Kripke's suggestion is to act as someone who ignores the meaning of "true" and wishes to learn it. Having no preconceptions about truth, thus, we take it momentarily to be "empty". Since we know how to evaluate arithmetical sentences, an obvious option is to try to evaluate the sentences obtained applying the truth predicate to them. If we can do that, we can re-iterate whatever procedure we employed, giving a semantic value to a larger collection of sentences, and so on, gradually "filling" the meaning of "true". This process ends when a point is reached where, reiterating the "filling" step, no more sentences are interpreted than those already interpreted.[6]

[3] For an analysis of the mathematical aspects of Kripke's theory, see [Fitting 1986].

[4] For Tarski-hierarchies see [Halbach 1996].

[5] Essentially weak Kleene logic, strong Kleene logic, and supervaluations..

[6] Some interesting remarks concerning the epistemic aspects and the algorithmic nature of Kripke's treatment of the truth predicate can be found in [Cantini 2002].

Arithmetical sentences are naturally divided in two sets by $\mathbb{N}$: the set $V_{\mathbb{N}}$ of arithmetical sentences validated by $\mathbb{N}$, and the set $F_{\mathbb{N}}$ of arithmetical sentences falsified by $\mathbb{N}$ (i.e. whose negation is validated by it). Having in mind that our goal is to make $T\ulcorner\varphi\urcorner$ and φ equivalent, a natural thought is that we should expand $V_{\mathbb{N}}$ to the set $V_{\mathbb{N}} \cup \{T\ulcorner\varphi\urcorner \in \mathscr{L}_T \mid \varphi \in V_{\mathbb{N}}\}$, and $F_{\mathbb{N}}$ to $F_{\mathbb{N}} \cup \{T\ulcorner\varphi\urcorner \in \mathscr{L}_T \mid \varphi \in F_{\mathbb{N}}\}$. Call the first set $(V_{\mathbb{N}})^+$ and the second set $(F_{\mathbb{N}})^-$. In other words, we validate all the arithmetical sentences that are true-in-$\mathbb{N}$ plus the sentences formalizing the claim that such sentences are true, and dually we falsify the arithmetical sentences that are false-in-$\mathbb{N}$ plus the sentences formalizing the claim that such sentences are true.

Suppose now that φ_0 and ψ_0 are both in $(V_{\mathbb{N}})^+$, i.e. they are both validated by the model we have built (now, however, φ_0 and ψ_0 may belong to $\mathscr{L}_T \setminus \mathscr{L}$). Under many understandings of the conjunction, $\varphi_0 \wedge \psi_0$ should be validated as well. Also, under some understandings of the negation, $\neg\varphi_0$ and $\neg\psi_0$ should be falsified too. Similar considerations go for the other connectives and quantifiers. In short, we close $(V_{\mathbb{N}})^+$ and $(F_{\mathbb{N}})^-$ under preservation of truth-in-the-model and falsity-in-the-model, for a chosen evaluation scheme for the logical vocabulary. I will only consider the strong Kleene evaluation scheme (see Definition 1.4.4), that is the more relevant one for the objectives of the present work.

Now, closing $(V_{\mathbb{N}})^+$ and $(F_{\mathbb{N}})^-$ under strong Kleene preservation of truth-in-the-model and strong Kleene preservation of falsity-in-the-model, we enlarge $(V_{\mathbb{N}})^+$ and $(F_{\mathbb{N}})^-$. Therefore, following the above idea about interpreting T over already validated or falsified sentences, we should also evaluate the application of T to the members of the new sets, in the same way as we did passing from $V_{\mathbb{N}}$ to $(V_{\mathbb{N}})^+$ and from $F_{\mathbb{N}}$ to $(F_{\mathbb{N}})^-$. This procedure generates two new sets $((V_{\mathbb{N}})^+)^+$ and $((F_{\mathbb{N}})^-)^-$, such that $(V_{\mathbb{N}})^+ \subseteq ((V_{\mathbb{N}})^+)^+$ and $(F_{\mathbb{N}})^- \subseteq ((F_{\mathbb{N}})^-)^-$,[7] that are then again enlarged as just explained, and again and again.

[7]The superscript $^+$ and $^-$ indicate the process of enlarging the collection of validated and falsified sentences. This process will be described more precisely in the next section.

This process is described by an inductive construction, which eventually reaches a halt (for cardinality reasons), i.e. a *fixed point*. Once we reach the fixed point we have a pair of sets such that, re-applying the process described above to it, we obtain that very pair again. So, the fixed point I have described informally is given by two sets $V_{\mathbb{N}}^{\infty}$ and $F_{\mathbb{N}}^{\infty}$ such that $(V_{\mathbb{N}}^{\infty})^{+} = V_{\mathbb{N}}^{\infty}$ and $(F_{\mathbb{N}}^{\infty})^{-} = F_{\mathbb{N}}^{\infty}$. There are many different fixed points of this kind of procedure, but in all of them a sentence φ is validated exactly if $T\ulcorner\varphi\urcorner$ is validated, and similarly for falsified sentences.

What about the liar and other paradoxical sentences? Are they validated or falsified? It depends. In a *consistent* fixed point of Kripke's construction, namely in a fixed point where the validated sentences are disjoint from the falsified ones, the liar sentence is neither in the set of validated sentences, nor in the set of the falsified ones. Consider the above fixed point, given by $V_{\mathbb{N}}^{\infty}$ and $F_{\mathbb{N}}^{\infty}$. The trick is that, even if the original division of $\mathscr{L}$-sentences was exhaustive (every $\mathscr{L}$-sentence is either validated or falsified), this is not so any longer when we move to $\mathscr{L}_T$. The liar, for example, is neither obtained applying the truth predicate to an arithmetical sentence or to a sentence obtained applying the truth predicate to an arithmetical sentence (and so on), nor it is the (strong Kleene) consequence of sentences previously validated or falsified. Therefore, it is just out of $V_{\mathbb{N}}^{\infty}$ and $F_{\mathbb{N}}^{\infty}$.

I will now turn to a formally precise presentation of Kripke's construction.

2.3 Kripke's theory: formal presentation

Here and in the following, I will be concerned solely with the version of Kripke's theory that employs the strong Kleene evaluation scheme to interpret connectives and quantifiers. Other evaluation schemes can be used in Kripke's construction, but they are far less relevant for the present work.[8]

[8] An analysis of the strong Kleene version of Kripke's theory can be found in [M. Kremer 1988] and in [McGee 1991], Chapter 4. A version of Kripke's semantics with an evaluation scheme that properly extends strong Kleene semantics is in [Halbach 2011], Chapter 15.1. Investigations on

Definition 2.3.1. (Kripke's construction for $\mathscr{L}_T$)

For any ordered pair $\langle S_1, S_2\rangle \subseteq \omega \times \omega$, define the pair $\langle (S_1)^+, (S_2)^-\rangle$ so that:

1. $n \in (S_1)^+$ if $n \in S_1$, or

   (i) n is an atomic arithmetical sentence φ and $\mathbb{N} \models \varphi$, or

   (ii) n is $\neg\varphi$ and $\varphi \in \mathscr{L}_T$ and $\varphi \in S_2$, or

   (iii) n is $\varphi \wedge \psi$ and $\varphi, \psi \in \mathscr{L}_T$ and $\varphi \in S_1$ and $\psi \in S_1$, or

   (iv) n is $\varphi \vee \psi$ and $\varphi, \psi \in \mathscr{L}_T$ and $\varphi \in S_1$ or $\psi \in S_1$, or

   (v) n is $\forall x \chi(x)$, $\forall x \chi(x) \in \mathscr{L}_T$, and for all $t \in CTER_{\mathscr{L}_T}$ $\chi(t) \in S_1$, or

   (vi) n is Tt and $t \in CTER_{\mathscr{L}_T}$ and $t = \ulcorner \chi \urcorner$ and $\chi \in \mathscr{L}_T$ and $\chi \in S_1$.

2. $n \in (S_2)^-$ if $n \in S_2$, or

   (i) n is an atomic arithmetical sentence φ and $\mathbb{N} \not\models \varphi$, or

   (ii) n is $\neg\varphi$ and $\varphi \in \mathscr{L}_T$ and $\varphi \in S_1$, or

   (iii) n is $\varphi \wedge \psi$ and $\varphi, \psi \in \mathscr{L}_T$ and $\varphi \in S_2$ or $\psi \in S_2$, or

   (iv) n is $\varphi \vee \psi$ and $\varphi, \psi \in \mathscr{L}_T$ and $\varphi \in S_2$ and $\psi \in S_2$, or

   (v) n is $\forall x \chi(x)$, $\forall x \chi(x) \in \mathscr{L}_T$ and for a $t \in CTER_{\mathscr{L}_T}$ $\chi(t) \in S_2$, or

   (vi) n is Tt and $t \in CTER_{\mathscr{L}_T}$ and either t does not denote the code of a $\mathscr{L}_T$-sentence, or $t = \ulcorner \chi \urcorner$ and $\chi \in \mathscr{L}_T$ and $\chi \in S_2$. ▲

We can associate with the above definition an operator on pairs of subsets of ω.

Definition 2.3.2. (Kripke's jump operator)

Let $\zeta_1(n, S_1, S_2)$ abbreviate the right-hand side of item 1 of Definition 2.3.1, and let $\zeta_2(n, S_1, S_2)$ abbreviate the right-hand side of item 2. $\zeta_1(n, S_1, S_2)$ and $\zeta_2(n, S_1, S_2)$

Kripke's theory with the weak Kleene evaluation scheme include [Cain and Damnjanovic 1991], while studies on Kripke's theory with supervaluations include [Burgess 1986], [Cantini 1990], [Meadows 2013], [Hansen 2015].

are positive elementary formulae, positive in S_1 and S_2. Define an operator acting on pairs of sets, $\Phi : \wp(\omega) \times \wp(\omega) \longmapsto \wp(\omega) \times \wp(\omega)$, as follows:

$$\Phi(\langle S_1, S_2 \rangle) := \langle \{n \in \omega \,|\, \zeta_1(n, S_1, S_2)\}, \{n \in \omega \,|\, \zeta_2(n, S_1, S_2)\} \rangle.$$

Let's call Φ the *Kripke jump*; I write $\Phi(S_1, S_2)$ as a shorthand for $\Phi(\langle S_1, S_2 \rangle)$. ▲

Lemma 2.3.3.

The Kripke jump Φ is inclusive and monotone.[9]

The jump Φ encodes the construction given in Definition 2.3.1, and the main objects of Kripke's theory are its fixed points, so I turn to them now.

For any $P, Q \subseteq \omega$, put

$$\mathfrak{I}_\Phi(P, Q) := \bigcup_{\alpha \in Ord} \Phi^\alpha(P, Q),$$

where $\Phi^\alpha(P, Q)$ is the α-th iteration of the application of Φ to the pair $\langle P, Q \rangle$, i.e. the α-th *stage* of the construction of $\mathfrak{I}_\Phi(P, Q)$. As customary, this will also be indicated as $\mathfrak{I}^\alpha_\Phi(P, Q)$. By well-known facts about inductive definitions, for any $P, Q \subseteq \omega$, exactly one set $\mathfrak{I}_\Phi(P, Q)$ exists, and it is a *fixed point* of Φ, namely:

$$\Phi(\mathfrak{I}_\Phi(P, Q)) = \mathfrak{I}_\Phi(P, Q).$$

Expressions of the form $\mathfrak{I}_\Phi(P, Q)$ will be called "Kripke fixed points" or "Kripke fixed-point pairs". For every $P, Q \subseteq \omega$, the pair $\mathfrak{I}_\Phi(P, Q)$ is defined by simultaneous induction, and is inductive in P and Q.[10]

[9] I.e. $\langle P, Q \rangle \subseteq \Phi(P, Q)$ (inclusiveness), and if $P \subseteq P'$ and $Q \subseteq Q'$, then $\Phi(P, Q) \subseteq \Phi(P', Q')$ (monotonicity). For the facts about inductive definitions used here, see [Moschovakis 1974], Chapter 1.

[10] Notice that the fact that $\mathfrak{I}_\Phi(P, Q)$ is inductive in P and Q does not conflict with the negative clauses "$\mathbb{N} \not\models \varphi$" or "$t \notin SENT_{\mathscr{L}_T}$" of Definition 2.3.1 (items 2.(i) and 2.(vi) respectively), since $\{\varphi \in \mathscr{L} \,|\, \mathbb{N} \models \varphi\}$ and $SENT_{\mathscr{L}_T}$ are hyperelementary sets.

A particularly important Kripke fixed point is the following:

$$\mathfrak{I}_\Phi := \mathfrak{I}_\Phi(\emptyset,\emptyset) := \bigcup_{\alpha \in Ord} \Phi^\alpha(\emptyset,\emptyset).$$

$\mathfrak{I}_\Phi$ is the fixed point that was informally described at the end of Section 2.2. Let $\leq_S$ denote the partial ordering of tuples of sets induced by $\subseteq$, namely such that $\langle P_1,\ldots,P_n\rangle \leq_S \langle Q_1,\ldots Q_n\rangle$ if and only if ($P_1 \subseteq Q_1$ and $\ldots$ and $P_n \subseteq Q_n$). It is easy to see that $\mathfrak{I}_\Phi$ is the $\leq_S$-*least* element of the lattice of the fixed points of Φ, and thus it is called *the least Kripke fixed point*.[11] In fact, for every $P,Q \subseteq \omega$:

$$\mathfrak{I}_\Phi = \bigcap_{P,Q \subseteq \omega} \mathfrak{I}_\Phi(P,Q) \subseteq \mathfrak{I}_\Phi(P,Q).$$

In general, $\mathfrak{I}_\Phi(P,Q)$ is the fixed point of Φ obtained by running Kripke's construction starting with $\langle P,Q\rangle$. Kripke's construction can be done over every choice of P and Q. In fact, rather than a single set-theoretic construction, Kripke's theory gives us a format to build various fixed points.

To discuss Kripke's theory, it will be useful to distinguish between the two sets which constitute a Kripke fixed-point pair.

Definition 2.3.4. (Kripke truth-sets and Kripke falsity-sets)
The first element of a Kripke fixed point $\mathfrak{I}_\Phi(P,Q)$ is its *Kripke truth-set*, in symbols $\mathfrak{E}_\Phi(P,Q)$, the second element is its *Kripke falsity-set*, indicated by $\mathfrak{A}_\Phi(P,Q)$. In other words, $\mathfrak{I}_\Phi(P,Q) = \langle \mathfrak{E}_\Phi(P,Q), \mathfrak{A}_\Phi(P,Q)\rangle$. Let $\mathfrak{E}^\alpha_\Phi(P,Q)$ denote the α-th stage of the construction of $\mathfrak{E}_\Phi(P,Q)$, i.e. the first component of the α-th stage of the construction of the pair $\mathfrak{I}_\Phi(P,Q)$, and similarly for $\mathfrak{A}^\alpha_\Phi(P,Q)$. Note that $\mathfrak{E}_\Phi(P,Q)$ and $\mathfrak{A}_\Phi(P,Q)$ are inductive in P and Q. ▲

[11] In my presentation of Kripke's theory the fixed points of Φ form a *lattice* (immediate by Definition 2.3.2, Lemma 2.3.3, and the Knaster-Tarski Theorem), whereas in Kripke's original formulation they only form a *partial order*.

A fixed point $\mathfrak{I}_\Phi(P,Q)$ is called *consistent* if $\mathfrak{E}_\Phi(P,Q) \cap \mathfrak{A}_\Phi(P,Q) = \emptyset$, and *inconsistent* otherwise. Although inconsistent fixed points were not the focus of Kripke's original discourse, they are easily seen to exist: an immediate example is $\mathfrak{I}_\Phi(\omega,\omega)$. Inconsistent Kripke fixed points are not interesting *per se*, since they give inconsistent information about $\mathscr{L}_T$-sentences, some of which are both validated and falsified. Nevertheless, they are useful to check whether a certain sentence can be consistently validated or falsified: if φ is such that, for every $P,Q \subseteq \omega$, if $\varphi \in \mathfrak{E}_\Phi(P,Q)$ then $\mathfrak{I}_\Phi(P,Q)$ is inconsistent, then φ cannot be validated in Kripke's construction – similarly for falsified sentences.[12]

Kripke's jump is defined by making explicit the clauses for preservation of values 1 and 0 of the strong Kleene scheme, plus weak naïveté. We can express this fact associating an *evaluation* with every Kripke fixed point.

Definition 2.3.5. (Kripke evaluations)

For every $P,Q \subseteq \omega$, define the *Kripke evaluation generated by* $\mathfrak{I}_\Phi(P,Q)$ as the set $\mathscr{K}_{P,Q} \subseteq SENT_{\mathscr{L}_T} \times \{1,0\}$ such that:

$$\langle \varphi, \mathbf{v} \rangle \in \mathscr{K}_{P,Q} \text{ if and only if } \begin{cases} \varphi \in \mathfrak{E}_\Phi(P,Q) \text{ and } \mathbf{v} \text{ is } 1, \text{ or} \\ \varphi \in \mathfrak{A}_\Phi(P,Q) \text{ and } \mathbf{v} \text{ is } 0. \end{cases}$$

Note that $\mathscr{K}_{P,Q}$ is a *function* if and only if $\mathfrak{I}_\Phi(P,Q)$ is consistent. In those cases, I will write $\mathscr{K}_{P,Q}(\varphi) = 1$ or $\mathscr{K}_{P,Q}(\varphi) = 0$ without risk of confusion. ▲

2.4 Properties of Kripke's construction

One of the most notable features of Kripke's theory is that it respects weak naïveté, in the sense of the following result.

[12] In this work, I do not consider views that accepts inconsistencies, such as *dialethism* (see, e.g., [Priest 2006]). Kripke categorizes several kinds of fixed points and uses his classification to analyze various kinds of $\mathscr{L}_T$-sentences. For an analysis of Kripke's classification, see [Burgess 1986], pp. 666-670 and [Cantini 1996], pp. 203-207.

Theorem 2.4.1. (Kripke)

For every $\varphi \in \mathscr{L}_T$ and $P, Q \subseteq \omega$:

1. $\varphi \in \mathfrak{E}_\Phi(P,Q)$ *if and only if* $T\ulcorner\varphi\urcorner \in \mathfrak{E}_\Phi(P,Q)$.

2. $\varphi \in \mathfrak{A}_\Phi(P,Q)$ *if and only if* $T\ulcorner\varphi\urcorner \in \mathfrak{A}_\Phi(P,Q)$.

Proof.

By the fixed-point property of Φ, we have that

$$\varphi \in \mathfrak{E}_\Phi(P,Q) \text{ if and only if } T\ulcorner\varphi\urcorner \in (\mathfrak{E}_\Phi(P,Q))^+,$$

but $\mathfrak{E}_\Phi(P,Q)$ is the truth-set of a Kripke fixed point, so $\mathfrak{E}_\Phi(P,Q) = (\mathfrak{E}_\Phi(P,Q))^+$. The same reasoning gives us the case of $\mathfrak{A}_\Phi(P,Q)$. □

Kripke's theory recovers also a version of intersubstitutivity of truth.

Corollary 2.4.2. (Kripke/Field)

For every $\varphi, \psi, \chi \in \mathscr{L}_T$ and $P, Q \subseteq \omega$, if ψ and χ are alike except that ψ has an occurrence of φ of $\mathscr{L}_T$ where χ has an occurrence of $T\ulcorner\varphi\urcorner$, then:

- $\psi \in \mathfrak{E}_\Phi(P,Q)$ *if and only if* $\chi \in \mathfrak{E}_\Phi(P,Q)$.

- $\psi \in \mathfrak{A}_\Phi(P,Q)$ *if and only if* $\chi \in \mathfrak{A}_\Phi(P,Q)$.

Proof. [13]

Let φ be a sub-sentence of ψ. Let's define the *i-th occurrence of φ within ψ*, indicated as φ^i, as the i-th element of some non-repeating enumeration of the instances of φ occurring in ψ (say, from the left-hand side).

Let's define the *depth of the embedding of φ^i in ψ*, call it $d(\varphi^i, \psi)$, as a function from the set of occurrences of φ in ψ to ω as follows:

[13] This demonstration follows the strategy sketched in [Field 2008], p. 62. I give some details of the proof as I am not aware of any other place in the literature where a proof of this result is actually written down; moreover, this demonstration will be useful for some results later on in this work.

- $d(\varphi^i, \psi) = 0$, if ψ is φ.

- $d(\varphi^i, \psi) = 1$, if ψ is $\neg\varphi$, or $\varphi \wedge \sigma$, or $\sigma \wedge \varphi$, or $\varphi \vee \sigma$, or $\sigma \vee \varphi$ and the shown φ is φ^i, or if φ is $\sigma(t)$ for some closed $\mathscr{L}_T$-term t and ψ is $\forall x\sigma(x)$.

- $d(\varphi^i, \psi) = n+1$, if ψ is $\neg\sigma^\dagger$, or $\sigma^\dagger \wedge \sigma$, or $\sigma \wedge \sigma^\dagger$, or $\sigma^\dagger \vee \sigma$, or $\sigma \vee \sigma^\dagger$ and φ^i is in the shown $\sigma^\dagger$ and $d(\varphi^i, \sigma^\dagger) = n$, or ψ is $\forall x\sigma(x)$ and for every closed $\mathscr{L}_T$-term t, $d(\varphi^i, \sigma(t)) = n$.

Note that for every occurrence φ^i of the sentence φ in ψ there is exactly one $n \in \omega$ such that $d(\varphi^i, \psi) = n$. An easy induction shows that, for ψ and χ as in the statement of the corollary, the depth of embedding of φ^i in ψ is exactly the depth of embedding of the occurrence of $T\ulcorner\varphi\urcorner$ that replaces φ^i in ψ.

If $d(\varphi^i, \psi) = d(T\ulcorner\varphi\urcorner^i, \chi) = 0$, the result is immediate by Theorem 2.4.1.

If $d(\varphi^i, \psi) = 1$, the result follows by a proof by cases. I will only do one case (the universal quantifier case) for Kripke truth-sets, as the cases for the other connectives and for Kripke falsity-sets are similar. I will omit the sign $^+$, using the fixed point property of Kripke truth-sets. Let $\psi[\varphi^i/\sigma]_N$ indicate the non-uniform substitution of the i-th occurrence of φ in ψ with the sentence σ.[14]

$$\begin{aligned}
\forall x\varphi(x) \in \mathfrak{E}_\Phi(P,Q) &\text{ iff } \varphi(t) \in \mathfrak{E}_\Phi(P,Q) \text{ for all } t \in CTER_{\mathscr{L}_T} \\
&\text{ iff } T\ulcorner\varphi(t)\urcorner \in \mathfrak{E}_\Phi(P,Q) \text{ for all } t \in CTER_{\mathscr{L}_T} \text{ (Thm 2.4.1)} \\
&\text{ iff } \forall x T\ulcorner\varphi(\dot{x})\urcorner \in \mathfrak{E}_\Phi(P,Q).
\end{aligned}$$

But then $\forall x\varphi(x)[\varphi(t)^i/T\ulcorner\varphi(t)\urcorner]_N \in \mathfrak{E}_\Phi(P,Q)$ iff $\forall x T\ulcorner\varphi(\dot{x})\urcorner[T\ulcorner\varphi(t)\urcorner^i/\varphi(t)]_N \in \mathfrak{E}_\Phi(P,Q)$, for every closed $\mathscr{L}_T$-term t, as desired.[15]

[14] As noted in [Smiley 1962], non-uniform substitutions can be defined via uniform substitutions, even though doing it can be quite cumbersome. I don't give any detail here for the sake of brevity.

[15] Substitutions are a bit tricky for quantifiers. The statement of the corollary implies the possibility of substituting only (some instances of) different sub-sentences of a sentence, but this is not always possible. All the (countably infinitely many) sub-sentences occurring in $\forall x\sigma(x)$ at depth 1 have the same logical form, namely $\sigma(t)$. There is no universally quantified sentence s.t. some of its sub-

The case for $d(\varphi^i, \psi) = n+1$ follows by an identical proof, using as inductive hypothesis the claim up to $d(\varphi^i, \psi) = n$. Again, I will do one case as example.

$$\begin{aligned}
\forall x \sigma(x) \in \mathfrak{E}_\Phi(P,Q) & \text{ iff } \sigma(t) \in \mathfrak{E}_\Phi(P,Q) \text{ for all } t \in CTER_{\mathscr{L}_T} \\
& \text{ iff } \sigma(t)[\varphi^i / T\ulcorner\varphi\urcorner]_N \in \mathfrak{E}_\Phi(P,Q) \text{ for all } t \in CTER_{\mathscr{L}_T} \text{ (IH)} \\
& \text{ iff } \forall x \sigma(x)[\varphi^i / T\ulcorner\varphi\urcorner]_N \in \mathfrak{E}_\Phi(P,Q)
\end{aligned}$$

As the claims hold for arbitrary occurrences and sentences, the result follows. $\square$

Finally, let me show the algebraic reading of the connectives and quantifiers adopted in Kripke's theory, which follows strong Kleene semantics.

Lemma 2.4.3.

Let $P, Q \subseteq \omega$ be such that $\mathfrak{I}_\Phi(P,Q)$ is consistent. Then, for every $\varphi, \psi \in \mathscr{L}_T$, and every $\chi(x) \in FOR_{\mathscr{L}_T}$, we have that:[16]

$$\mathscr{K}_{P,Q}(\neg\varphi) = \begin{cases} 1 - \mathscr{K}_{P,Q}(\varphi), \textit{ if it is defined}, \\ \textit{undefined, otherwise.} \end{cases}$$

$$\mathscr{K}_{P,Q}(\varphi \wedge \psi) = \begin{cases} 0, \textit{ if } \mathscr{K}_{P,Q}(\varphi) = 0 \textit{ or } \mathscr{K}_{P,Q}(\psi) = 0, \\ \min(\mathscr{K}_{P,Q}(\varphi), \mathscr{K}_{P,Q}(\psi)), \textit{ if they are both defined}, \\ \textit{undefined, otherwise.} \end{cases}$$

$$\mathscr{K}_{P,Q}(\varphi \vee \psi) = \begin{cases} 1, \textit{ if } \mathscr{K}_{P,Q}(\varphi) = 1 \textit{ or } \mathscr{K}_{P,Q}(\psi) = 1, \\ \max(\mathscr{K}_{P,Q}(\varphi), \mathscr{K}_{P,Q}(\psi)), \textit{ if they are both defined}, \\ \textit{undefined, otherwise.} \end{cases}$$

---

sentences occurring at depth 1 have the form $\sigma(t)$ while others have the form $T\ulcorner\sigma(t)\urcorner$, for $\sigma(t)$ distinct from $T\ulcorner\sigma(t)\urcorner$. Contrast this with, e.g., conjunction: we can take $\sigma \wedge \chi$ and add T to only one of its sub-sentences occurring at depth 1, e.g. $T\ulcorner\sigma\urcorner \wedge \chi$. This is why, in the proof, I replace every sub-sentence occurring at depth 1 in a universally quantified sentence.

[16] As recalled in Definition 2.3.5, if $\mathfrak{I}_\Phi(P,Q)$ is consistent, then $\mathscr{K}_{P,Q}$ is a function.

$$\mathscr{K}_{P,Q}(\forall x\chi(x)) = \begin{cases} 0,\ \textit{if}\ \mathscr{K}_{P,Q}(\chi(t_n)) = 0\ \textit{for some}\ t_n \in TER_{\mathscr{L}_T}, \\ \inf\{\mathscr{K}_{P,Q}(\chi(t_n))\,|\,t_n \in TER_{\mathscr{L}_T}\},\ \textit{if they are all defined}, \\ \textit{undefined, otherwise.} \end{cases}$$

2.5 Difficulties for Kripke's theory

Despite its success in accommodating weak naïveté and intersubstitutivity of truth, and the interest of Kripke fixed points, Kripke's theory has been subject to multiple criticisms for some truth-theoretic and logical shortcomings.

2.5.1 Gaps, paradoxes, and expressive limitations

Let me start reviewing the difficulties of Kripke's theory from its handling of the liar sentence (and relevantly similar cases).

Lemma 2.5.1.
For every $P,Q \subseteq \omega$, if $\lambda \in \mathfrak{E}_\Phi(P,Q) \cup \mathfrak{A}_\Phi(P,Q)$, then $\mathfrak{I}_\Phi(P,Q)$ is inconsistent.

Proof.
Let $P,Q \subseteq \omega$. If $\lambda \in \mathfrak{E}_\Phi(P,Q)$, then $\neg\lambda \in \mathfrak{A}_\Phi(P,Q)$ by the fixed-point property of $\mathfrak{I}_\Phi(P,Q)$. Moreover, by the definition of λ, we have $\neg T\ulcorner\lambda\urcorner \in \mathfrak{E}_\Phi(P,Q)$ and, by Corollary 2.4.2, $\neg\lambda \in \mathfrak{E}_\Phi(P,Q)$, showing $\mathfrak{I}_\Phi(P,Q)$ to be inconsistent. The proof for the case when $\lambda \in \mathfrak{A}_\Phi(P,Q)$ is dual. □

Kripke's construction yields ⟨Kripke truth-set, Kripke falsity-set⟩ pairs, and λ cannot be in any element of these pairs, unless consistency is violated. Kripke's construction cannot consistently situate the liar sentence in any of the fixed points it generates. But Kripke fixed points are, essentially, the models of $\mathscr{L}_T$-sentences that Kripke's theory produces. So, Kripke's theory has no way of interpreting the liar sentence in a consistent way.[17]

Sentences that are neither in $\mathfrak{E}_\Phi(P,Q)$ nor in $\mathfrak{A}_\Phi(P,Q)$ have come to be called "gappy" with respect to $\mathfrak{I}_\Phi(P,Q)$. Let's give a precise definition of this notion.

Definition 2.5.2. (The C-gap of a Kripke fixed point)
For every $P,Q \subseteq \omega$, the *C-gap of* $\mathfrak{I}_\Phi(P,Q)$ is $SENT_{\mathscr{L}_T} \setminus (\mathfrak{E}_\Phi(P,Q) \cup \mathfrak{A}_\Phi(P,Q))$. Call the elements of a C-gap C-gappy (with respect to a given Kripke fixed point). The name "C-gaps" reflects the fact that they are defined using the *complement* of the union of truth-set and falsity-set of Kripke fixed points. ▲

C-gaps are central in many discussions of Kripke's theory, although their role and importance are not uncontroversial. It is sometimes argued that, thanks to C-gaps, Kripke's theory *treats successfully* paradoxical sentences such as the liar. Now, as we have seen, Kripke's theory produces consistent fixed points that vindicate some forms of weak naïveté and intersubstitutivity of truth,[18] while they avoid the problems potentially arising from the liar and similar sentences by safely relegating them to C-gaps.[19] It is, however, improper to say that Kripke's theory provides a treatment of paradoxical sentences. It is not the case that λ *is C-gappy according to Kripke's theory*, as there is no way for Kripke's theory to talk about C-gaps. Since Kripke's construction only yields pairs of the form ⟨Kripke truth-set, Kripke falsity-set⟩, clearly it does not also give their complement. But the problem is deeper than that: the C-gaps of consistent fixed points are not sets that Kripke's theory can define, as Kripke's construction employs an inductive definition, but the former are not inductive sets.

[17] If φ is neither in $\mathfrak{E}_\Phi(P,Q)$ nor in $\mathfrak{A}_\Phi(P,Q)$, this does *not* mean that the claim that φ is not true in the sense of the truth predicate is validated in this fixed point, namely that $\neg T\ulcorner\varphi\urcorner \in \mathfrak{E}_\Phi(P,Q)$. For otherwise $\neg\varphi$ would be in $\mathfrak{E}_\Phi(P,Q)$ by Corollary 2.4.2, and φ would be in $\mathfrak{A}_\Phi(P,Q)$.

[18] See Theorem 2.4.1 and Corollary 2.4.2.

[19] Lemma 2.5.1.

Lemma 2.5.3. [20]

For every $P,Q \subseteq \omega$, if $\mathfrak{I}_\Phi(P,Q)$ is consistent, $SENT_{\mathscr{L}_T} \setminus (\mathfrak{E}_\Phi(P,Q) \cup \mathfrak{A}_\Phi(P,Q))$ is co-inductive and not inductive in P,Q.

If two sets $A,B \subseteq \omega$ form a consistent Kripke fixed-point pair, then $\langle A,B\rangle = \mathfrak{I}_\Phi(P,Q)$ for some $P,Q \subseteq \omega$, but there is no inductive definition over P and Q that defines the resulting C-gap. The latter set cannot result from any application of Kripke's construction and no information on membership in it can be obtained from the positive inductive means available to Kripke's theory. C-gaps are only visible "from the outside" of Kripke's theory, as it were, and the strictly Kripkean theorist cannot but be silent about them, since talking about C-gaps requires to accept resources that go essentially beyond Kripke's theory.

The extent and radicalness of such silence are well emphasized by Field:

> For [C-gappy] sentences, we can't say that they're true, or that they aren't, or that they're false, or that they aren't. We can't say whether or not they are "[C-]gappy" [...]. And our inability to say these things can't be attributed to ignorance, for we don't accept that there is a truth about the matter. This isn't to say that we think there is no truth about the matter: we don't think there is, and we don't think there isn't. And we don't think there either is or isn't. Paracompleteness runs deep. ([Field 2008], p. 72)

This observation rectifies a noteworthy misconception about Kripke's theory, namely that C-gappy sentences (with respect to a consistent Kripke fixed point) are assigned the semantic value $1/2$ of strong Kleene semantics *by Kripke's construction*. There are certainly good reasons for wanting this association. C-gappy sentences are closed under the strong Kleene clauses for the preservation of value $1/2$, just like the sentences in a Kripke truth-set are closed under the strong Kleene clauses for the preservation of 1. Moreover, some readings of the semantic value

[20] This result is folklore: for a proof see, for example [McGee 1991], Corollary 5.11, p. 113.

$1/2$ of the strong Kleene semantics, such as *undefined*, seem pertinent to the categorization of sentences that fail to come out as 1 or 0 in a consistent Kripke fixed point, and perhaps also to sentences typically considered paradoxical. However, as Kripke's theory is in general unable to say anything at all about C-gappy sentences, it is also unable to assign them a semantic value.

In fact, as I gave Kripke's inductive construction without even mentioning C-gappy sentences, I defined Kripke evaluations without even mentioning value $1/2$: the Kripke evaluation generated by a consistent fixed point is an *essentially partial assignment* of values 1 and 0 to $\mathscr{L}_T$-sentences, under which some sentences receive just no value.[21] Kripke himself warned his readers against this confusion:

> I have been amazed to hear my use of the Kleene valuation compared occasionally to the proposals of those who favor [...] positing extra truth values beyond truth and falsity [...]. "Undefined" is not an *extra* [...] value anymore than – in Kleene's book – **u** is an extra *number* in sec. 63. ([Kripke 1975], pp. 700-701, footnote 18)[22]

The association of Kripke's theory with strong Kleene semantics, thus, is not incorrect, but misguided some interpreters. For example, consider the following passage by Tim Maudlin:

> It is unclear to me on exactly what grounds Kripke denies that his undefined is a [...] value. The analogy to Kleene's **u** is weak in the

[21] That a consistent Kripke fixed point generates a partial and not total function as its Kripke evaluation is immediate from Lemma 2.5.1.

[22] The reference is to [Kleene 1952], pp. 323-332 (where **u** is used for $1/2$). This part of Kripke's quote seems uncontroversial and clear. However, Kripke goes on with more contentious claims:

> Nor should it be said that "classical logic" does not generally hold [...]. *If* certain sentences express propositions, any tautological truth-function of them expresses a true proposition. Of course formulas, even with the form of tautologies, which have components that do not express propositions may have [value-]functions that do not express propositions either. [...] conventions for handling sentences that do not express propositions are not in any philosophically significant sense "changes in logic".

The claim that Kripke's theory does not abandon "classical logic" seems difficult to maintain: every classical law has instances that are not in the truth-set of any consistent Kripke fixed point (e.g. $\lambda \vee \neg\lambda$, immediate from Lemma 2.5.1). I do not wish to enter this debate here; however, I will advance a possible explanation for Kripke's attitude towards logic in Subsection 4.2.3.

> following way: there is a large collection of mathematical functions and relations whose domain ought to include all numbers. We expect that numbers can be added, multiplied, divided by one another, and so on. [...]. But **u** seems not to be admissible to the domain of these operations. [...]. But the functions of [...] values seem to be much more restricted: in semantics, the [...] values of sentences function largely to determine the [...] values of sentences of which they are immediate semantic constituents. So, if **u** is added to the truth tables of the logical particles, and a clear account is given of the conditions under which a sentence is assigned **u**, it is unclear what else needs to be done to render **u** a [...] value. ([Maudlin 2004], p. 50, footnote 6)

Maudlin is right in that one *can* consider $1/2$ (or **u**, as he calls it) as a semantic value. But it is clearly not the case that "Kripke's undefined" *is* a value for Kripke's theory, exactly as C-gaps are not sets given by Kripke's theory. Maudlin conflates Kripke's theory and the theory consisting of Kripke fixed points *and their C-gaps*.

Availing oneself Kripke fixed points and their C-gaps, one can assign a value to the sentences that fail to have a value in the former, thus making this failure explicit. The next definition formalizes this idea.

Definition 2.5.4. (Kripke C-gappy evaluations)

For every $P,Q \subseteq \omega$, define the *Kripke C-gappy evaluation generated by* $\mathfrak{I}_\Phi(P,Q)$ as the set $\mathscr{K}^+_{P,Q} \subseteq SENT_{\mathscr{L}_T} \times \{1,0,1/2\}$ such that:

$$\langle \varphi, \mathbf{v}\rangle \in \mathscr{K}^+_{P,Q} \text{ if and only if } \begin{cases} \varphi \in \mathfrak{E}_\Phi(P,Q) \text{ and } \mathbf{v} \text{ is } 1, \text{ or} \\ \varphi \in \mathfrak{A}_\Phi(P,Q) \text{ and } \mathbf{v} \text{ is } 0, \text{ or} \\ \varphi \in SENT_{\mathscr{L}_T} \setminus (\mathfrak{E}_\Phi(P,Q) \cup \mathfrak{A}_\Phi(P,Q)) \text{ and } \mathbf{v} \text{ is } 1/2 \end{cases}$$

As for the old Kripke evaluations, also the C-gappy ones are functions exactly when generated by a consistent Kripke fixed point. ▲

It is easy to expand Lemma 2.4.3 to the case of Kripke C-gappy evaluations.

Lemma 2.5.5.

Let $P,Q \subseteq \omega$ be such that $\mathfrak{I}_\Phi(P,Q)$ is consistent. Then, for every $\varphi, \psi \in \mathscr{L}_T$, and *every $\chi(x) \in FOR_{\mathscr{L}_T}$, we have that:*

$$\mathscr{K}^+_{P,Q}(\neg\varphi) = 1 - \mathscr{K}^+_{P,Q}.$$
$$\mathscr{K}^+_{P,Q}(\varphi \wedge \psi) = \min(\mathscr{K}^+_{P,Q}(\varphi), \mathscr{K}^+_{P,Q}(\psi)).$$
$$\mathscr{K}^+_{P,Q}(\varphi \vee \psi) = \max(\mathscr{K}^+_{P,Q}(\varphi), \mathscr{K}^+_{P,Q}(\psi)).$$
$$\mathscr{K}^+_{P,Q}(\forall x\chi(x)) = \inf\{\mathscr{K}^+_{P,Q}(\chi(t_n)) \,|\, t_n \in TER_{\mathscr{L}_T}\}.$$

Definition 2.5.4, together with Lemma 2.5.3, makes it clear that, within the framework of Kripke's construction, evaluating the sentences that are C-gappy with respect to a consistent Kripke fixed point requires means than exceed those provided by Kripke's theory. The liar is one such sentence, as Lemma 2.5.1 shows it to be C-gappy with respect to every consistent fixed point. This remark will prove crucial to devise improvements to Kripke's theory.[23]

Opinions diverge a lot on the importance of C-gaps, both *per se* and within Kripke's theory. On the one hand, for example, [Soames 1999] argues that C-gaps are a major result of Kripke's construction: they constitute a new conception of gappiness, alternative to the Russellian and to the Strawsonian paradigms, that can be used to develop a new idea of partially defined predicates, and that may afford us an account of truth which improves substantially on the Tarskian one.[24] On the other hand, [McGee 1989] calls C-gaps a "red herring", as "they are not essential either to understanding the mathematical construction Kripke develops or to obtaining a correct understanding of truth".[25] I will enter this debate only tangentially. My main aim will be to argue that, even if we enrich Kripke's theory considering all the C-gaps generated by Kripke fixed point, despite the flexibility

[23]See Subsection 4.5.1.
[24]See especially pp. 170-176 and pp. 180-181.
[25]Page 530.

and theoretical fruitfulness of the resulting construction, we cannot provide an ultimately satisfactory account of the paradoxes. *Prima facie*, it would seem that one could use C-gaps to treat problematic cases such as the liar. In reality, however, the apparatus of Kripke's theory cannot *express* appropriately the status of several sentences typically considered paradoxical or somehow problematic (I will focus on the liar and the truth-teller for definiteness, but the following considerations apply to many other sentences).

A fundamental feature of the liar seems to be that it *cannot* be given a classical value. Under rather minimal assumptions (T obeying weak naïveté and $\neg$ working as strong Kleene negation for at least 1 and 0), it is impossible to assign a classical semantic value to the liar. How are we going to express this fact? The formalization of the sentence "the liar sentence is not true and not false" (taking the falsity of φ as the truth of $\neg\varphi$) is a non-starter, since assuming intersubstitutivity of truth and double negation elimination, if we accept $\neg T\ulcorner\lambda\urcorner \wedge \neg T\ulcorner\neg\lambda\urcorner$, then we accept $\lambda \wedge \neg T\ulcorner\neg\neg T\ulcorner\lambda\urcorner\urcorner$ and $\lambda \wedge \lambda$.

The following biconditional gives a more interesting option:

(G) The liar sentence is true exactly if its negation is true

(G) voices a natural reaction to the liar and it is informative because it makes explicit the fundamental feature from above, exhibiting nicely the behavior of the liar sentence in the classical derivations of the paradox.

The formalization of (G), however, cannot be validated by Kripke's theory, whether extended with C-gaps or not.

Corollary 2.5.6. [26]
For every $P,Q \subseteq \omega$, if $T\ulcorner\lambda\urcorner \equiv T\ulcorner\neg\lambda\urcorner \in \mathfrak{E}_\Phi(P,Q)$, then $\mathfrak{I}_\Phi(P,Q)$ is inconsistent.

[26] The proof is immediate from the proof of Lemma 2.5.1.

Moreover, it is not the case that there is exactly one consistent Kripke fixed point that has in its C-gap all the sentences that share the fundamental feature highlighted above. So, the best that Kripke's theory can do in order to account for such feature is to characterize the sentences that share it in the meta-theory used to construct Kripke fixed points, via properties of the set of fixed points. Here is how Kripke does it: "a sentence is *paradoxical* if it has no truth value in any [consistent] fixed point".[27] This solution, however, is not very satisfactory. The models for $\mathscr{L}_T$-sentences that result from Kripke's construction are Kripke fixed points, and the above observation tells us that Kripke's theory gives us no single model that accounts for the fact that the liar cannot have a classical value. In order to grasp this fact, we have to consider all the consistent Kripke fixed points, which are uncountably infinitely many. But one of the main purposes of devising semantic theories of truth is to come up with *one* good model that we accept, as we accept its interpretation of the object-language sentences and its account of their properties.

This kind of problem reappears also in less catastrophic cases. For example, in virtually every semantic theory the truth-teller *can* consistently have any semantic value available. In Kripke's framework, it can have value 1, 0, or $1/2$. This fundamental feature of the truth-teller, however, is not easy to account for. First, it is unclear that we can come up with a sentence that expresses this feature and that can be formalized in a language such as $\mathscr{L}_T$ (i.e. comparable to the role played by (G) for the liar). Second, the semantic value of the truth-teller in any single fixed point is not of great help. The truth-teller is in the C-gap of the least Kripke fixed point, but this is not very telling: also the liar sentence is in the C-gap of the least Kripke fixed point, but the two sentences seem to have very different, indeed almost opposite, fundamental features. The oscillation between 1, 0, and $1/2$ in consistent fixed points is sometimes taken as a characterization of the nature of τ, but this is

[27] [Kripke 1975], p. 708.

a poor characterization indeed. If one takes values 1 and 0 seriously (for example as indicating conditions for acceptance of a sentence, or of its negation), she could argue that Kripke's theory treats τ incoherently: can we really take seriously the values given to τ, when both $\mathfrak{I}_\Phi(\{\tau\},\emptyset)$ and $\mathfrak{I}_\Phi(\emptyset,\{\tau\})$ are consistent and agree on the values of sentences not involving τ?[28] Of course, the advocate of Kripke's theory could reply that Kripke's construction allows us to say something about its peculiarity precisely because of this oscillation, and to contrast it with, say, the liar, by considering the collection of consistent Kripke fixed points. This is a fair point but, as above, it resorts to several models. No single model for $\mathscr{L}_T$ built via Kripke's theory can capture and convey the most telling fact about the nature of the truth-teller, i.e. that we can give it whatever semantic value we like.

Some Kripke fixed points do express crucial features of some $\mathscr{L}_T$-sentences, thus characterizing them via a single model: e.g., the greatest intrinsic fixed point "is the unique "largest" interpretation of $T(x)$ which is consistent with our intuitive idea of truth and makes no arbitrary choices in truth assignment".[29] Still, for many interesting kinds of paradoxes, we have seen, the Kripkean treatment is highly inadequate. To improve on it, we would like to evaluate liar-like sentences, truth-teller-like sentences and so on in some meaningful way that expresses their status, while preserving the values of $\mathfrak{I}_\Phi$. It would be nice to have *one* evaluation that either validates formalized sentences that express interesting features of paradoxes (e.g. a formalization of (G) for the liar), or that assigns to paradoxical cases some semantic values that codify their fundamental features. However, it is quite unclear how this could be accomplished within Kripke's framework.[30]

[28]The informally characterized "sentences not involving τ" can be defined precisely as the set of sentences not *grounded in* $\mathscr{L}\cup\{\tau\}$, for Kripke's notion of grounding (see [Kripke 1975], p. 694). Kripke's notion of grounding will be discussed and expanded in Subsection 5.1.1.

[29][Kripke 1975], pp. 709-710. The notion of intrinsic fixed point is also explained there.

[30]In Chapter 5, I will construct a model that, as argued in Subsection 6.3.2, addresses these issues in a satisfactory way. One might object to my criticism, arguing that it is reasonable to analyze the features of liar, truth-teller, and other cases via several models, by analogy with possible world semantics for modal claims, since such features were described in modal terms. This objection,

There are also different kinds of claims one may want to make about typically paradoxical sentences that are also not justifiable on the basis of Kripke's theory and its C-gaps. For example, Corollary 2.5.6 would also hold replacing λ with $\kappa_\supset$. For every Kripke fixed point $\mathfrak{I}_\Phi(P,Q)$, a negated sentence $\neg\varphi$ is in the truth-set (falsity-set, C-gap) of $\mathfrak{I}_\Phi(P,Q)$ if and only if so is $\varphi \supset \bot$. Therefore, not only there are interesting relations between $\neg$ and $\supset$, but also between the liar and Curry's paradox. For example, one may want to express the following idea:

(LC) The liar sentence is equivalent to Curry's sentence.

(LC) may be used (perhaps together with other claims) to argue that the liar and Curry's paradox are just two superficially different versions of the same antinomy. Whatever its merits, this conclusion cannot be advanced on the basis of Kripke's theory plus its C-gaps, as the sentence $\lambda \equiv \kappa_\supset$, that formalizes it, is in the C-gap of every consistent Kripke fixed point ($\kappa_\supset$ is the formalization of Curry's sentence available in $\mathscr{L}_T$), even if λ and $\kappa_\supset$ always get the same value in Kripke's theory.

2.5.2 Weakness of the logic

The foregoing observations point already at another problem of Kripke's theory, that has been widely criticized, namely the "weakness of the logic", closely related to the adoption of the strong Kleene scheme.

To begin with, in every consistent Kripke fixed point, every logical law has at

however, is not compelling, since: (i) we are not considering languages featuring truth-theoretic and modal vocabulary, and (ii) it is questionable that the above features should be understood via some independent notion of modality, as they seem simply to express the possibility or impossibility *of some evaluations* for $\mathscr{L}_T$-sentences – it is not the case that, say, the truth-teller is "possible" in some independently interesting sense of possibility. It is clear, then, that a single model for a $\mathscr{L}_T$ that can express the above features of liar-like cases, truth-teller-like cases and so on has a decisive expressive advantage over Kripke's theory on the account of truth-theoretic paradoxes. For investigations on truth and modalities, and on the related topic of predicate approaches to modality, see [Horsten 1998], [Halbach 2002], [Horsten 2003], [Halbach, Leitgeb, and Welch 2005], [Leitgeb 2006], [Halbach 2006], [Halbach and Welch 2009].

least one instance that does not get value 1 in it. This fact, occasionally sloganeered as "Kripke's theory has no laws", is a consequence of the adoption of the strong Kleene scheme, which does not validate any logical law. This may not be worrying if the rejected laws are controversial for truth. As recalled in Section 1.4, Horsten emphasizes that everyone accepting excluded middle accepts its instance $\lambda \vee \neg\lambda$, and thus also $(\lambda \wedge \neg T\ulcorner\lambda\urcorner) \vee (\neg\lambda \wedge T\ulcorner\lambda\urcorner)$. But $(\lambda \wedge \neg T\ulcorner\lambda\urcorner) \vee (\neg\lambda \wedge T\ulcorner\lambda\urcorner)$ is undesirable if the predicate T has to resemble a coherent concept of truth, as it shows that either we accept something untrue, or that we deny something true. Nevertheless, in Kripke's theory also desirable principles are ruled out, such as $\neg(\varphi \wedge \neg\varphi)$. Moreover, the fact that the only conditional we have at our disposal is the strong Kleene conditional makes it difficult to formalize some claims in an uncontroversial way. For example, if we wanted to formalize the apparently harmless "if φ, then φ" within Kripke's theory, we would resort to $\varphi \supset \varphi$, which is equivalent to the highly controversial excluded middle $\varphi \vee \neg\varphi$. Similar considerations go for the formalization of "if everything φs, then in particular t φs", and so on.

The logical weaknesses of Kripke's theory have a crucial impact on its account of naïve truth. In fact, the lack of logical laws makes it impossible for Kripke's theory to validate every Tarski biconditional – formalized in the only way available to Kripke's theory, namely via the strong Kleene biconditional. If there were a consistent Kripke fixed point $\mathfrak{I}_\Phi(P,Q)$ having $T\ulcorner\lambda\urcorner \equiv \lambda$ in its truth-set, it would also have in its truth-set $\lambda \equiv \lambda$ and $\neg\lambda \vee \lambda$, which is shown to be impossible by Lemma 2.5.1 and its proof. Therefore, everyone who attaches a crucial importance to strong naïveté, and thus to recovering some version of Tarski biconditionals in a semantic theory of truth, will be obviously dissatisfied with Kripke's theory.

Another crucial point about the "logic" of Kripke's theory concerns the introduction and elimination rules for connectives and quantifiers.[31] Semantic in-

[31] I am concerned with *semantic inferences* rather than inference rules in some formal system. I will, however, stick to the term "rule", since this use should not cause any confusion.

ferences are often expressed via some notion of logical consequence, and there are several options to associate a logical consequence with Kripke's construction, quantifying over various sets of Kripke fixed points (e.g. all the consistent ones, all the intrinsic ones, etc.). The following definition, I take it, is general enough to cover all the interesting cases, it is natural for its structural similarity with the classical notion of logical consequence, and it imposes rather modest constraints.

Definition 2.5.7. (Logical consequence for Kripke's theory)
For every non-empty set $\mathcal{A}$ of consistent Kripke fixed points, define the *Kripke $\mathcal{A}$-consequence*, in symbols $\models_K^{\mathcal{A}}$, as follows:

$$\Gamma \models_K^{\mathcal{A}} \psi \text{ if and only if for every } \mathfrak{I}_\Phi(P,Q) \in \mathcal{A},$$
$$\text{if, for every} \varphi \in \Gamma, \varphi \in \mathfrak{E}_\Phi(P,Q), \text{ then } \psi \in \mathfrak{E}_\Phi(P,Q).$$

▲

Remark 2.5.8.
Definition 2.5.7 has the only cost of excluding sets of fixed points that include inconsistent elements. The motivation for this restriction is quite straightforward. Let $\models_K^{\mathcal{B}}$ be the logical consequence relation defined as in Definition 2.5.7 but with the difference that $\mathcal{B}$ is an arbitrary set of Kripke fixed points. In this variant, many basic pieces of semantic reasoning are lost. For example, it is easy to see that, for some such set $\mathcal{B}$, there are sentences $\varphi, \psi \in \mathscr{L}_T$ s.t. $\models_K^{\mathcal{B}} \varphi \supset \psi$ but $\varphi \not\models_K^{\mathcal{B}} \psi$, which upsets a fundamental relation between basically every conditional (including the strong Kleene material conditional), and logical consequence. It is possible to liberalize Definition 2.5.7 to include *some* inconsistent Kripke fixed points: e.g., the problem just mentioned would be avoided restricting the variant of Definition 2.5.7 to sets $\mathcal{B}$ including consistent fixed points or inconsistent ones that are closed under explosion. Since we cannot use arbitrary Kripke fixed points, however, it

seems reasonable to ignore the inconsistent ones altogether: after all, even the most "harmless" inconsistent fixed points are inconsistent, and thus they do not yield interpretations of $\mathscr{L}_T$-sentences that we would accept.[32,33] ▲

Using this notion of logical consequence, it is easily seen that the classical rules to introduce and eliminate $\wedge$, $\vee$, and $\forall$ are recovered. For example, for every Kripke $\mathcal{A}$-consequence $\models^{\mathcal{A}}_K$, we have:

$$\varphi \wedge \psi \models^{\mathcal{A}}_K \varphi \text{ and } \varphi \wedge \psi \models^{\mathcal{A}}_K \psi,$$

$$\varphi, \psi \models^{\mathcal{A}}_K \varphi \wedge \psi.$$

Also, in every Kripke $\mathcal{A}$-consequence $\models^{\mathcal{A}}_K$, the rules for $\neg$ are the rules of strong Kleene negation, which approximate closely the classical rules for negation.[34]

The naïve rules for truth are also validated, as for every Kripke $\mathcal{A}$-consequence $\models^{\mathcal{A}}_K$ and every $\mathscr{L}_T$-sentence φ we have:

$$\varphi \models^{\mathcal{A}}_K T\ulcorner\varphi\urcorner, \qquad T\ulcorner\varphi\urcorner \models^{\mathcal{A}}_K \varphi.$$

The rules for $\supset$ are more problematic: while *modus ponens* is recovered, as

$$\varphi, \varphi \supset \psi \models^{\mathcal{A}}_K \psi,$$

conditional introduction poses some difficulty. The classical rule for conditional

[32] Recall that here I do not consider perspectives that accept inconsistencies, such as dialethism.

[33] There are, of course, still alternative ways of associating a notion of logical consequence with Kripke's theory. The structure of the logical consequence of Definition 2.5.7, however, makes it the closest (that I know of) to the standard notion of (semantic) logical consequence familiar from classical first-order logic, since it quantifies over the models of $\mathscr{L}_T$ afforded by Kripke's theory, namely Kripke fixed points, and it uses them just as classical models are used in the classical consequence relation – this may become more evident if one reformulates Definition 2.5.7 via Kripke evaluation functions. I take that, in absence of good motivations to do so, we should not deviate from the classical, well-established conceptual picture of logical consequence.

[34] See [Urquhart 2002], p. 259.

introduction is often stated as follows (for the classical logical consequence $\models$):

$$\text{if } \varphi \models \psi, \text{ then } \models \varphi \supset \psi.$$

If Kripke $\mathcal{A}$-consequence is an appropriate account of (semantic) logical consequence for Kripke's theory, as it seems to be, then classical conditional introduction is out of reach in Kripke's framework. For every set of consistent Kripke fixed points $\mathcal{A}$, in fact, $\lambda \models_K^{\mathcal{A}} \lambda$, but by Lemma 2.5.1 $\not\models_K^{\mathcal{A}} \lambda \supset \lambda$.[35]

Kripke's theory does not give us the classical conditional introduction, as the conditional available to Kripke's theory is the strong Kleene conditional. Since $\varphi \supset \varphi$ is just $\neg\varphi \vee \varphi$, the treatment of conditional sentences is influenced by the failure of excluded middle. The absence of the *classical* conditional introduction, however, is not at all deplorable in the context of naïve truth. As [Field 2008] rightly stresses, because of Curry's paradox, if a theory licenses intersubstitutivity of truth, then *modus ponens* and the classical introduction of the conditional cannot hold together for any conditional.[36,37]

This fact introduces a conceptually fundamental difference between what Field calls "conditional assertion" and what he calls "assertion of a conditional". *Conditional assertion* indicates that a sentence ψ follows from a sentence φ (plus possibly other sentences) via the notion of logical consequence of our theory (in the case

[35]Stricter notions of logical consequence, i.e. such that less sentences enter the logical consequence relation, will not help: the meta-theory of Kripke's theory is classical, and whenever we use a classically defined notion of logical consequence (as in Definition 2.5.7), λ is a logical consequence of λ, reproducing the above problem. At most, a stricter notion of logical consequence may reduce the semantic inferences validated by the theory: for example, in the logical consequence relation discussed in [Halbach and Horsten 2006] (see pp. 691-692) and associated with the formal system PKF (given in a sequent calculus formalism), not only the classical conditional introduction, but also *modus ponens* is invalidated.

[36]See the derivations of Curry's paradox in [Field 2008] p. 84 or p. 281. An option that I do not consider, in order to take consistently together *modus ponens* and classical introduction of the conditional, is altering the *structural* properties of the logical consequence relation, particularly transitivity and contraction. Axiomatic substructural approaches to naïve truth and truth-theoretic paradoxes have been developed extensively in recent years by Zardini (see his [2011], [2013]).

[37]I am also not considering the option to drop the rule *modus ponens*, as it happens in some dialethic theories, such as Priest's LP (see his [1979], [2006]).

of Kripke's $\mathcal{A}$-consequence, $\Gamma, \varphi \models^{\mathcal{A}}_{K} \psi$ is a conditional assertion); the *assertion of a conditional*, instead, is given when a conditional sentence follows logically, via the notion of logical consequence of our theory, from a set of sentences, possibly empty (e.g. $\Gamma \models^{\mathcal{A}}_{K} \varphi \supset \psi$, for Kripke's $\mathcal{A}$-consequence). These two notions, Field claims, have to come apart in the presence of naïve truth. The relation between these two notions, one might add, should be investigated, in order to have a full understanding of the effects of the adoption of naïve truth.

Since we cannot have the classical conditional introduction, if we are not willing to give *modus ponens* up, a natural question comes up: what weakening of the classical conditional introduction can we get? As for Kripke's theory, there is a certain consensus that we can introduce a conditional whenever the excluded middle holds for its antecedent.[38] In terms of Kripke $\mathcal{A}$-consequence, we can express this idea thus:

$$\text{if } \varphi \models^{\mathcal{A}}_{K} \psi \text{ and } \models^{\mathcal{A}}_{K} \varphi \vee \neg\varphi, \text{ then } \models^{\mathcal{A}}_{K} \varphi \supset \psi. \tag{2.1}$$

However, this rule is very uninteresting, due to the weakness of the strong Kleene conditional. Essentially, we can introduce a conditional in classical contexts, namely in the case of typically unproblematic inferences. But we are left in the dark for the inferences concerning non-classical cases, usually more problematic and difficult to accommodate. In fact, we cannot express via the conditional $\supset$ the logical consequence relations holding between many interesting sentences (typically paradoxical) that fail to satisfy the excluded middle condition in any collection of consistent Kripke fixed points. For example, for every Kripke $\mathcal{A}$-consequence $\models^{\mathcal{A}}_{K}$, we have $\lambda \models^{\mathcal{A}}_{K} \neg\lambda$ and $\neg\lambda \models^{\mathcal{A}}_{K} \lambda$, but clearly no $\supset$-conditional holds between λ and $\neg\lambda$ (compare with the discussion on sentence (G) above).

[38] This was done (using the condition $T\ulcorner\varphi\urcorner \vee T\ulcorner\neg\varphi\urcorner$ rather than LEM proper), for a proof-theoretic notion of logical consequence, in [Horsten 2011] (pp. 132-133), in the context of a natural deduction version of the axiomatic theory PKF.

The use of the strong Kleene conditional suggests that there is little hope to improve on the unsatisfactory condition (2.1) within Kripke's framework.[39] And this is a serious flaw of Kripke's approach. One of the main aims of theories of truth is to bring the truth-theoretic discourse at the object-language level. It would be nice to bring as many inferences as possible to the object-language level, especially including inferences concerning paradoxes, as they stand in need of explanation. A conditional introduction that is strong, well-motivated, and conceptually interesting is fundamental for this purpose. The dim prospects of getting such conditional introduction in Kripke's theory, thus, limit substantially the interest of this theory.

2.6 Field's program

The difficulties of Kripke's theory and of the resulting account of the semantics of $\mathscr{L}_T$ explored in the previous section can be summarized as follows:

(K1) In Kripke's theory (i.e. consistent Kripke fixed points) no logical laws are validated, and there is no prospect of recovering an interesting and strong semantic rule governing the introduction of the conditional.

(K2) For every sentence φ and every consistent Kripke fixed point $\mathfrak{I}_\Phi(P,Q)$, φ has value 1 (0, $1/2$) in $\mathfrak{I}_\Phi(P,Q)$ if and only if $T\ulcorner\varphi\urcorner$ has value 1 (0, $1/2$) in it. However, this fact cannot be represented within the theory via a biconditional in many cases, including apparently simple ones – inconsistent fixed points are an exception, but they are not interesting for my purposes.

(K3) Kripke's theory fails to deliver a satisfactory account of paradoxes. On the one hand, Kripke's official theory does not include any treatment of the liar, the truth-teller, and other paradoxes. Moreover, some central features of

[39] This is accepted by several commentators, because of the strong Kleene scheme. In fact, it seems difficult to improve on (2.1) even liberalizing my (already quite liberal) definition of Kripke $\mathcal{A}$-consequence including some kinds of inconsistent fixed points.

these paradoxes cannot be addressed even enlarging Kripke's theory to include C-gaps, as the discussion in Subsection 2.5.1 shows.

Hartry Field has argued convincingly that several difficulties of Kripke's theory can be successfully tackled by *adding a primitive conditional to the language*, and giving it a interpretation radically different from the strong Kleene conditional. Looking at my analysis of Kripke's theory and its problems, summarized in **(K1)**-**(K3)**, it seems clear that a conditional more interesting that Kripke's $\supset$ would help, even if some of the difficulties would not be automatically addressed, such as the ones concerning truth-teller-like cases.

(K1) would be addressed by a new conditional for which some classical laws hold (together with naïveté). Moreover, every Kripke $\mathcal{A}$-consequence $\models^{\mathcal{A}}_{K}$ recovers strong Kleene validities but, just like the strong Kleene consequence relation, lacks their law-version: e.g. $\varphi \wedge \psi \models^{\mathcal{A}}_{K} \varphi$, but the $\supset$-law is invalid: $\not\models^{\mathcal{A}}_{K} \varphi \wedge \psi \supset \varphi$. That *many of the instances* of the desired conditional principles are not validated (including instances of Tarski biconditionals) can also be seen as a consequence of the lack of a conditional having a strong introduction rule. Since we cannot improve on this situation with the strong Kleene conditional, an obvious option is to introduce a new conditional for which some interesting laws hold, or that validates *modus ponens* and a nice introduction rule, or both.

As to **(K2)**, it is clear that we cannot improve on the amount of Tarski biconditionals that Kripke's theory can formalize and validate using the resources of strong Kleene semantics. However, it may be possible to add a primitive conditional to $\mathscr{L}_T$ and use it to validate more Tarski biconditionals (formalized with the new conditional), perhaps even all of them.

As to **(K3)**, finally, not only a new semantics is required to improve on the (lacking or poor) treatment of paradoxes in Kripke's theory, but a new conditional may be used to account for some aspects of paradoxes, such as those indicated by statements (G) and (LC), which remain unexpressed in Kripke's framework.

Field's arguments for the need of supplementing Kripke's theory with a new conditional, and his work to accomplish this goal, contributed to promote a line of research that partly motivates the present work. Although Field himself, to my knowledge, never presented this line of investigation as a research agenda, I think it is interesting to articulate it as a systematic program. Here is my rendering of it.

FIELD'S PROGRAM The search for models that preserve Kripke's theory and add to it a *well-behaved* conditional.

Now, it is clear that the conditional we find in Kripke's theory is not well-behaved, as it is subject to all the difficulties emphasized previously. A conditional which improves on the strong Kleene one in terms of overcoming **(K1)**-**(K3)** is surely "better-behaved". However, it is natural to ask more general questions: what should we expect from a well-behaved conditional (especially in the context of paracomplete theories of naïve truth)? Are there plausible criteria that we should impose on conditionals which are candidates to address FIELD'S PROGRAM? More generally, when is a connective or a quantifier well-behaved?

2.7 Well-behaved connectives and quantifiers

I know of no systematic attempt to answer the questions that close the previous section, either in Field's work, or even in the context of paracomplete theories of naïve truth more in general. For this reason, in this section, I advance a proposal for the notion of well-behaved connective or quantifier.

Undoubtedly, there is a lot of latitude in the idea of well-behaved connective or quantifier. Still, the question of how to interpret connectives and quantifiers is pressing in the context of theories of naïve truth. Such theories, in fact, need to revise classical semantics, and this forces us to inquire into the interpretations of the logical vocabulary that support (some form of) naïveté: if this investigation has to have some philosophical significance, it should provide an interesting and well-justified reading of the logical vocabulary. Thanks to the success of Kripke's theory in preserving a good account of negation, conjunction, disjunction, and of the universal quantifier, my main concern will be a well-behaved conditional. However, it is preferable to consider the behavior the connectives and quantifiers as a whole, also to grasp the role and the specificity of the conditional.

Here, I do not try to cover all the options that one can consider to tell whether some connectives and quantifiers are well-behaved; rather, I propose some general criteria that may be broadly acceptable and theoretically fruitful, at least within the domain of theories of truth.[40] Let C be the set of connectives and quantifiers of the language we are considering, call such language $\mathscr{L}_T^C$. Of course, I hold fixed that our language includes a predicate for truth, that should be given a naïve interpretation. Let's say that a semantics for $\mathscr{L}_T^C$ that gives a naïve interpretation to T *makes C well-behaved* if the following conditions are met.

(W1) Every element of C has a clear and simple *meaning*. For example, it should be possible to describe the meaning of every $\circ$ in C via some simple algebraic rule (given by the semantics), or to axiomatize adequately the $\mathscr{L}_T^C$-sentences validated by the semantics,[41] or to characterize every $\circ$ in C via suitable introduction and elimination rules (validated by the semantics).

(W2) Every $\circ$ in C as interpreted by the semantics should be closed under introduction and elimination rules that approximate the classical ones.

(W3) Conferring a meaning on the connectives and quantifiers in C, as specified in **(W1)**, should allow us to render (as much as possible) the intuitions we

[40]For the following requirements, I have in mind connectives and first-order quantifiers of the kind that can be found in classical logic and in its generalizations.

[41]The idea of *adequate* axiomatization of theories of truth will be discussed in Subsection 3.5.3.

aimed at capturing with the development of our semantic theory.

The criteria expressed by **(W1)**-**(W3)** are quite vague, but they are sufficiently intelligible to serve as a regulative ideal to guide our efforts in providing a satisfactory semantics, and as a touchstone to test our results.

The main condition is clearly **(W1)**. A requirement along the lines of **(W1)** should be desirable for anyone who attributes a substantial role to formal semantics in casting light on notions as important as those expressed by connectives and quantifiers. In order for our semantic theory to be a philosophically interesting story about the languages we consider, the effects of naïveté etc., **(W1)** requires that the theory not only provides technical results, but also that it could be used to reach an understanding of the key elements of first-order languages, i.e. connectives and quantifiers. No particular restriction is imposed on what counts as "meaning" for them, accepting accounts that range from model-theoretic/algebraic accounts, to axiomatizations, to accounts via introduction and elimination rules.

The thought underlying **(W2)** is that, ideally, we should interpret $\mathscr{L}_T^C$ detaching ourselves as little as possible from the classical way of using connectives and quantifiers. Classical rules are taken as standard because nearly everybody agrees that a theory of naïve truth should provide a classical interpretation of the truth-free part of the language, especially since we use mathematical or syntactic languages as base languages. This requirement makes a lot of sense for paracomplete theories: Kripke's theory shows that it is easy to give a paracomplete theory that respects the classical introduction and elimination rules for $\wedge$, $\vee$, $\forall$. **(W2)** specifies that, in interpreting another connective or quantifier that has a classical *pendant* (such as a negation or a conditional stronger than the strong Kleene ones), we should try to make its introduction and elimination rules close to the classical rules, disrupting classical semantics as little as possible. I introduce no precise notion of *closeness* to classical rules, but this vagueness does not make **(W2)** unintelligible.

The main target of **(W2)** will be conditional introduction. While Curry's paradox dictates that we cannot hope to recover the classical conditional introduction (accepting *modus ponens* and intersubstitutivity of truth), I argued above that the rule more plausibly associated with Kripke's theory (namely (2.1)) is rather uninteresting. We would like to have a new conditional $\rightarrow^*$ obeying a weakening of classical conditional introduction which is interesting and as strong as possible. Let a new conditional $\rightarrow^*$ be in the set C of our connectives and quantifiers, and suppose that our semantics provides us a notion of logical consequence $\models^*$. A proposal to realize this intuition is to aim at a rule of the following kind:

$$\Gamma, \varphi \models^* \psi \text{ and } \Gamma \models^* \mathcal{H}(\varphi, \psi) \quad \text{if and only if} \quad \Gamma \models^* \varphi \rightarrow^* \psi,$$

where φ, ψ are $\mathscr{L}_T^C$-sentences, Γ is a set of $\mathscr{L}_T^C$-sentences, and $\mathcal{H}(\varphi, \psi)$ is an $\mathscr{L}_T^C$-sentence indicating a condition on φ, ψ, or both. To make such a rule interesting, we should require that $\mathcal{H}(\varphi, \psi)$ does not include conditions that are too strong and imply $\Gamma \models^* \varphi \rightarrow^* \psi$ by themselves. Naïve truth makes it a bit cumbersome to formulate this requirement, which is given in the following definition.

Definition 2.7.1. (Genuine conditions for conditional introduction)
Define the following sets by transfinite recursion, for Δ a set of $\mathscr{L}_T^C$-sentences.

$$\varphi \in \mathcal{S}(\Delta) \text{ iff } \begin{cases} \varphi \text{ is a subsentence of some } \psi \in \Delta, \text{ or} \\ \varphi \text{ results from a subsentence of some } \psi \in \Delta \text{ replacing} \\ \quad \text{one occurrence of a subsentence } T\ulcorner\chi\urcorner \text{ of } \psi \text{ with } \chi \end{cases}$$

$$\mathcal{S}^\alpha(\Delta) := \mathcal{S}\Big(\bigcup_{\beta<\alpha} \mathcal{S}^\beta(\Delta)\Big), \qquad \mathcal{S}^\infty(\Delta) := \bigcup_{\alpha \in Ord} \mathcal{S}^\alpha(\Delta).$$

By a small notational abuse, if $\Delta = \{\varphi\}$, I will write $\mathcal{S}^\infty(\varphi)$. So, $\mathcal{S}^\infty(\varphi)$ contains all the sentences obtained considering iteratively the subsentences of φ and the results of truth-theoretic disquotation applied to them (recall that every sentence is a subsentence of itself) – similarly for $\mathcal{S}^\infty(\Delta)$.

Let's say that a schematic $\mathscr{L}_T^C$-sentence $\mathcal{H}(\varphi,\psi)$ is a *genuine condition for conditional introduction* for the logical consequence $\models^*$ if, for every $\varphi,\psi \in \mathscr{L}_T^C$ and $\Gamma \subseteq \mathscr{L}_T^C$, the following holds:

(G-C-Intro) $\quad \Gamma,\varphi \models^* \psi$ and $\Gamma \models^* \mathcal{H}(\varphi,\psi)$ if and only if $\Gamma \models^* \varphi \to^* \psi$,

where:

1. $\varphi \in \mathcal{S}^\infty(\mathcal{H}(\varphi,\psi))$, or $\psi \in \mathcal{S}^\infty(\mathcal{H}(\varphi,\psi))$, or both, and
2. there is no $\chi \in \mathcal{S}^\infty(\mathcal{H}(\varphi,\psi))$ s.t. for every $\varphi_0,\psi_0 \in \mathscr{L}_T^C$:

$$\Gamma \models^* \varphi_0 \to^* \psi_0 \text{ if and only if } \Gamma \models^* \chi_0$$

   where χ_0 results from $\mathcal{S}^\infty(\mathcal{H}(\varphi,\psi))$ instead of χ, putting φ_0 instead of φ and ψ_0 instead of ψ.

In short, $\mathcal{H}(\varphi,\psi)$ is a genuine condition for conditional introduction if it is about φ, ψ, or both (item 1), and neither $\mathcal{H}(\varphi,\psi)$ nor any of its subschemata (possibly buried into instances of T) are equivalent to the $\models^*$-validity of all the corresponding conditionals, making the condition $\mathcal{H}(\varphi,\psi)$ uninteresting (item 2). ▲

This definition filters out undesirable rules, such as the following one:

$$\Gamma,\varphi \models_K^{\mathcal{A}} \psi \text{ and } \Gamma \models_K^{\mathcal{A}} (\varphi \vee \neg\varphi) \vee T\ulcorner \neg\varphi \vee \psi \urcorner \text{ if and only if } \Gamma \models_K^{\mathcal{A}} \varphi \supset \psi.$$

This rule is clearly uninformative, since it tells us that, whenever $\varphi \models_K^{\mathcal{A}} \psi$, we can introduce the corresponding conditional if φ has a classical value, or ... the whole conditional holds! Of course, a similar condition is not acceptable to express what a conditional assertion has to satisfy in order to become the assertion of a conditional, as it already contains the condition that the corresponding conditional is validated. This example makes it also clear why we need to consider subschemata, possibly buried into the scope of instances of T.

Kripke $\mathcal{A}$-consequence admits a genuine condition for conditional introduction, namely taking $\mathcal{H}(\varphi, \psi)$ to be $(\varphi \vee \neg\varphi) \vee \psi$.[42] However, this condition is just a weakening of (2.1), and it is as conceptually unsatisfactory as the latter.

In general, a genuine condition for conditional-introduction is philosophically crucial to understand the relation between conditional assertion and the assertion of a conditional. Since conditional assertions are weaker than assertions of a conditional (the former follow always from the latter, but not *vice versa*), a genuine condition for conditional introduction is decisive to determine in what the difference between these two notions consists, and when we can turn a reasoning under assumptions into a conditional fully expressible in the object-language.

(W3) concerns the philosophical import of semantic theories of naïve truth. Such theories are often developed to capture intuitions about truth, which then combine with intuitions about the connectives, the quantifiers, the ground language, and so on. It may also be that, in developing models to capture these intuitions, new features of the languages under considerations are discovered, which then interact with our pre-formal philosophical views. A criterion such as **(W3)** aims at tightening the link between our philosophical motivations and the results of the formal modeling, requiring that the modeling confers a meaning to the connectives and quantifiers which accords as much as possible with the informal intuitions we aimed at capturing – for whatever notion of meaning we choose.

An advantage of requirements such as **(W1)**-**(W3)** is that, even if they are con-

[42]The proof that $(\varphi \vee \neg\varphi) \vee \psi$ yields a genuine condition is immediate; to see that it is not equivalent to the Kripke $\mathcal{A}$-validity of all the corresponding $\supset$-conditionals, for any Kripke $\mathcal{A}$-consequence, consider $\mathcal{H}(0 = 0, \lambda)$.

ceived and developed in the context of paracomplete theories, they can easily be shared by theorists advocating completely different approaches. It is also clear that these criteria are not in any way superfluous or philosophically inert, as FIELD'S PROGRAM can be addressed in a purely technical fashion, without the methodologically principled approach I advocate here.

2.8 Summary of the chapter

In this chapter, I presented Kripke's theory, I gave a critical analysis of it, and I articulated a research program, that I called FIELD'S PROGRAM.

In Section 2.1, I gave an overview of the context in which Kripke's theory emerged and of the importance of Kripke's contribution. In Section 2.2, I provided an informal sketch of Kripke's theory, followed by the formal presentation of Section 2.3, and by a review of some facts about Kripke's theory in Section 2.4. In Section 2.5, I developed a critical analysis of Kripke's theory, trying to give an original discussion of its truth-theoretic shortcomings (Subsection 2.5.1) and of its logical weaknesses (Subsection 2.5.2). Section 2.6 presented and motivated FIELD'S PROGRAM. Finally, in Section 2.7 I explored the notion of well-behaved connective or quantifier, a key component of FIELD'S PROGRAM.

The next chapter is a critical analysis of *Field's revision theory*. Field's theory is a fundamental object of investigation for the present work, as it is one of the most advanced, successful, and interesting paracomplete theories of naïve truth currently available; moreover, it tackles many of the problems of Kripke's theory. Nevertheless, Field's theory presents also noteworthy difficulties, as we will see.

Chapter 3

Field's Theory

3.1 Introduction

In this chapter, I will present and discuss a theory of truth developed by Hartry Field in a recent series of works, especially [2003a], [2003b], [2004a], [2004b], [2004c], [2007], the monograph [Field 2008], and the comments in [Field 2011] discussing the observations in [D.A. Martin 2011] and [Welch 2011].

I will in speak of "Field's theory", meaning the model-theoretic construction that is expounded at length in [Field 2008]. Actually, the label "Field's theory" is not entirely justified, since Field developed two other theories in order to supplement Kripke's theory with a new conditional, namely the constructions in [Field 2002] and [Field 2014] respectively. Of course, for reasons of space it is not possible to discuss every theory of naïve truth with a good conditional on the market, not even restricting the attention to the paracomplete camp.[1] I will focus on Field's revision theory because it is, in the writer's opinion, the most developed and interesting construction proposed by Field, and one of the most advanced and successful paracomplete theories of naïve truth currently available. Moreover,

[1] A paracomplete theory of naïve truth with a conditional stronger than Field's is in [Bacon 2013a].

amongst all the constructions in the literature that I know of, Field's revision theory is the one that gets closer to addressing FIELD'S PROGRAM. In addition, most of the criticisms I will raise in conjunction to Field's revision theory are adaptable to Field's other theories, and I will occasionally refer to the latter on specific points. For these reasons, I will refer to Field's revision theory simply as "Field's theory", without fear of confusion.

3.1.1 Field's theory: some background

As hinted at the end of the previous chapter, Field's approach does not consist in giving a different interpretation to the material conditional of the language $\mathscr{L}_T$. Rather, the method followed by Field (and others) is to expand the language $\mathscr{L}_T$ to a new language that features a primitive binary connective $\rightarrow$, and to provide a model for the so-enlarged language where the new conditional $\rightarrow$ behaves "better" than the material conditional of Kripke's theory.

Definition 3.1.1. (The language $\mathscr{L}_T^{\rightarrow}$ and Curry's sentence)
Let $\mathscr{L}_T^{\rightarrow}$ be the first-order language obtained adding the new logical symbol "$\rightarrow$" to the connectives of $\mathscr{L}_T$. $\leftrightarrow$ is defined as usual, namely letting $\varphi \leftrightarrow \psi$ be a shorthand for $(\varphi \rightarrow \psi) \wedge (\psi \rightarrow \varphi)$. Fix some arithmetical coding of the expressions of the enlarged language $\mathscr{L}_T^{\rightarrow}$ as explained in Subsection 1.3. All the notational conventions in force for $\mathscr{L}_T$ will be adopted for $\mathscr{L}_T^{\rightarrow}$ as well. Finally, let κ, *Curry's sentence*, be the sentence $Tt_\kappa \rightarrow 0 = 1$ s.t. $t_\kappa = \ulcorner \kappa \urcorner = \ulcorner Tt_\kappa \rightarrow 0 = 1 \urcorner$. ▲

Addressing FIELD'S PROGRAM becomes now the task to devise a model for $\mathscr{L}_T^{\rightarrow}$ that respects Kripke's interpretation of $\neg$, $\wedge$, $\vee$, $\forall$, and of the truth predicate, which at the same time gives an interpretation to sentences of the form $\varphi \rightarrow \psi$ that makes $\rightarrow$ well-behaved. To avoid confusion with the material conditional, I will only refer to the new connective as "the conditional", calling a sentence "a conditional sentence", or "a conditional" for short, if it has the form $\varphi \rightarrow \chi$.

In order to keep Kripke's treatment of $\neg$, $\wedge$, $\vee$, $\forall$, and of the truth predicate, a common strategy is to make use of Kripke's construction *inside* other constructions, as it were, in ways which will be made precise in the course of this work. As we will see, this is also the strategy adopted in Field's theory. Using Kripke's theory inside Field's revision theory requires to run *Kripke's construction for the language* $\mathscr{L}_T^{\rightarrow}$, although Kripke's construction was originally developed for the language $\mathscr{L}_T$. Let me specify this point with the following definition.

Definition 3.1.2. (Kripke's construction for $\mathscr{L}_T^{\rightarrow}$)
By *Kripke's construction for* $\mathscr{L}_T^{\rightarrow}$, I mean the variant of Definition 2.3.1 that is obtained replacing all the references to the conditional-free language $\mathscr{L}_T$ with references to the the enlarged language $\mathscr{L}_T^{\rightarrow}$. ▲

So, for example, in Kripke's construction for $\mathscr{L}_T^{\rightarrow}$, elements of ω may code sentences of $\mathscr{L}_T^{\rightarrow}$ (and not of $\mathscr{L}_T$), and the instantiations of universally quantified sentences include all the terms of $\mathscr{L}_T^{\rightarrow}$ (and not only the terms of $\mathscr{L}_T$). However, there are no substantial differences between the two versions of Kripke's construction (i.e. for the language $\mathscr{L}_T$ and for $\mathscr{L}_T^{\rightarrow}$) as far as the interpretation of sentences is concerned. Neither version of Kripke's construction, in fact, has clauses for the evaluation of conditional sentences, that are simply ignored. There is no clause of Kripke's inductive definition that puts a conditional sentence $\varphi \rightarrow \psi$ into a Kripke truth-set or falsity-set in which $\varphi \rightarrow \psi$ was not already – of course the same holds for the applications of the truth predicate to conditional sentences. In short, Kripke's construction for $\mathscr{L}_T^{\rightarrow}$ treats conditionals as atomic sentences for which there is no evaluation clause. Of course, one can run Kripke's construction on sets that include conditional sentences already, in which case they will be carried on up to the fixed point.

Since I will deal mainly with the language $\mathscr{L}_T^{\rightarrow}$ in the remainder of this work, from now on I will speak simply of "Kripke's construction" meaning Kripke's construction for the language $\mathscr{L}_T^{\rightarrow}$, as specified in Definition 3.1.2, unless indicated otherwise. The same applies to all the related expressions, such as "Kripke fixed points", and so on. Note that all the observations made about Kripke's theory for the language $\mathscr{L}_T$ in the previous chapter carry over unchanged to Kripke's theory for the language $\mathscr{L}_T^{\rightarrow}$ since Kripke's theory features no treatment of the primitive conditional $\rightarrow$. For this reason, and to improve readability, I will not adopt any notational variant for Kripke's theory for $\mathscr{L}_T^{\rightarrow}$: for example, $\mathfrak{I}_\Phi(P,Q)$ is the Kripke fixed point obtained starting from the pair $\langle P,Q\rangle$, with the only difference that we are now always working with the language $\mathscr{L}_T^{\rightarrow}$. Similar considerations go for all the other pieces of notation I introduced in the previous chapter.

One may wonder why I did not formulate Kripke's theory for $\mathscr{L}_T^{\rightarrow}$ from the beginning, rather than formulating it for the language $\mathscr{L}_T$ in the previous chapter and then formulating it again here for $\mathscr{L}_T^{\rightarrow}$. Of course, that could have easily been done, and that would have avoided the need for the present re-formulation and the subsequent elucidations. However, such a choice would have been misconceived, for several reasons: (i) it would have been unfaithful to the original work by Kripke, and to its motivations and accomplishments; (ii) the distinction between semantics for $\mathscr{L}_T$ and semantics for $\mathscr{L}_T^{\rightarrow}$ is useful for expository purposes, e.g. to emphasize that the kind of investigation undertook in the present work requires to *add* a *new* conditional to Kripke's theory, and not to re-interpret the logical vocabulary that is already available there; moreover, formulating Kripke's theory for $\mathscr{L}_T^{\rightarrow}$ from the very start would have suggested the need for a new conditional *already before* analyzing Kripke's theory and the expressive problems related to its material conditional, but it is this analysis that motivates the search for a new conditional.[2] (iii) introducing $\mathscr{L}_T$ and $\mathscr{L}_T^{\rightarrow}$ separately, together with Kripke's construction for them, will help to present some models I will propose in Subsection 4.5.1.

[2]A similar methodological choice is followed in [Field 2008] (p. 83 and following), and the resulting exposition of Field's theory and its relations with Kripke's one is quite clear, in my opinion.

Finally, let me say something about the general approach adopted in Field's theory. Field's revision theory, as the way I call it suggests, is based on a *revision-theoretic* conception. The main difference between the usual revisionary approaches and Field's own is that in the former it is the truth predicate which undergoes revision, i.e. the value of sentences of the form $T\ulcorner\varphi\urcorner$ is revised, whereas in Field's theory it is the value of conditional sentences that is revised.[3]

3.1.2 Field's approach and FIELD'S PROGRAM

As far as I am aware, Field never specifies exactly what he requires from his new conditional, what would make it "well-behaved". Nonetheless, some of his observations may give us an idea of his criteria.

In introducing the prospect of adding a conditional to Kripke's theory, Field suggests that no monotone construction that validates Intersubstitutivity of Truth, such as Kripke's, can make the conditional "non-trivial":

> [... the] monotonicity principle rules out any *non-trivial conditional* "$\rightarrow$" *for which* "$\varphi \rightarrow \varphi$" *is a law*, at least if the system is non-dialetheic. ([Field 2008], p. 83, my emphasis)

Thus it seems that validating the schema $\varphi \rightarrow \varphi$ (i.e. conferring the/a designated value to all its instances) is an essential criterion to make $\rightarrow$ interesting.

In the presence of intersubstitutivity of truth, the schema $\varphi \rightarrow \varphi$ yields immediately the unrestricted Tarski schema $\varphi \leftrightarrow T\ulcorner\varphi\urcorner$ – accepting that $\varphi \rightarrow \varphi$ yields $\varphi \leftrightarrow \varphi$, of course. In fact, it seems that the intersubstitutivity of truth, the schema $\varphi \rightarrow \varphi$, and the Tarski schema are inextricably at the core of what Field considers an improvement on Kripke's theory via a well-behaved conditional:

[3] A classical source for revision-theoretic approaches to truth is [Belnap and Gupta 1993]. [Yablo 2003] advocates a kind of revision similar to Field's one, and the relations between Yablo's and Field's theories are discussed in [Field 2004c], pp. 69-74.

> Thus [with the development of Field's theory] I keep a promise made long ago, of getting a logic for the conditional strong enough so that the Intersubstitutivity Principle for truth follows from the Tarski schema as well as entailing it. ([Field 2008], p. 253)

Field, thus, attaches some importance at the idea that the Tarski schema should entail intersubstitutivity of truth. Clearly, however, while the Tarski schema follows immediately from $\varphi \leftrightarrow \varphi$ via intersubstitutivity, recovering the intersubstitutivity of truth from the Tarski schema requires several logical principles, many of which involve the conditional. The principles in the following list, for example, are recovered in Field's theory and guarantee that, if the Tarski schema is validated, so is intersubstitutivity of truth (I do not introduce a precise notion of logical consequence for Field's theory yet: for now, read "$\varphi \models_F \psi$" as "in every model given by Field's construction, if φ gets the designated value, so does ψ"):[4]

$$\begin{aligned}
\varphi \rightarrow \psi &\models_F \neg\psi \rightarrow \neg\varphi \\
\varphi \rightarrow \psi &\models_F (\varphi \wedge \chi) \rightarrow (\psi \wedge \chi) \\
\varphi \rightarrow \psi &\models_F (\chi \wedge \varphi) \rightarrow (\chi \wedge \psi) \\
\varphi \rightarrow \psi &\models_F (\psi \rightarrow \chi) \rightarrow (\varphi \wedge \chi) \\
\varphi \rightarrow \psi &\models_F (\chi \rightarrow \varphi) \rightarrow (\chi \wedge \psi) \\
\varphi(x) &\models_F \forall x \varphi(x) \\
\forall x(\varphi(x) \rightarrow \psi(x)) &\models_F \forall x \varphi(x) \rightarrow \forall x \psi(x)
\end{aligned}$$

Field, in fact, requires still more from the conditional. *Modus ponens*, for instance, figures prominently between the features of a good conditional for him. Discussing why Kripke's theory formulated with strong Kleene logic is preferable to other versions, he argues thus:

[4] $\models_F$ is discussed in Section 3.4. For the following principles, see [Field 2008] p. 253 and [Field 2003a] pp. 150-151 and endnote 9. $\models_F$ also obeys the classical structural rules.

> [...] its semantics provides a more natural basis for extending the theory to one that contains a *well-behaved* (modus-ponens-*obeying*) *conditional*. ([Field 2008], p. 66, fn. 13, my emphasis)

In another passage, Field attaches a crucial importance to *modus ponens*:

> The absence of a conditional with *modus ponens* is enough to *prevent any semblance of ordinary reasoning*. (Among many other things, it precludes reasoning by mathematical induction: a theory that invalidates *modus ponens* will clearly invalidate induction as well.) ([Field 2008], p. 369, my emphasis)

Now Field has required quite a lot of features from a good conditional, but they do not suffice yet to specify what would constitute a "decent" substitute for Kripke's material conditional. The latter, in fact, is a "totally inadequate" conditional in those settings where the excluded middle fails, since

> [...] with [the material conditional of Kripke's theory] as one's candidate for $\rightarrow$, one wouldn't even get such elementary laws of the conditional as $\varphi \rightarrow \varphi$, $\varphi \rightarrow (\varphi \vee \psi)$, or the inference from $\varphi \rightarrow \psi$ to $(\chi \rightarrow \varphi) \rightarrow (\chi \rightarrow \psi)$. ... The lack of a conditional (and also of a biconditional) cripples ordinary reasoning. ([Field 2008], p. 73)

This small anthology of passages suffices to show that Field is interested in recovering some collection of laws and rules that have the following features: they can be used to characterize naïve truth (e.g. via intersubstitutivity of truth), or they allow us to retain some intuitively desirable forms of reasoning (e.g. mathematical induction), or they seem to have an independent interest (e.g. $\varphi \rightarrow \varphi$, $\varphi \rightarrow \varphi \vee \psi$).

However, Field never specifies a collection of target principles, nor a clear and unique rationale for them. This methodological choice has mixed consequences for his approach. On the one hand, it grants a nice status to his theory. In fact, Field's theory adds to our knowledge of the semantics of naïve truth by showing that some principles which were not known to be consistent with naïve truth are consistent with it. This makes Field's revision theory instructive and useful, and Field can take himself to be advancing the knowledge about naïve truth, without pretending to give *the right semantics* for it. If another theory is found which recovers more principles than Field's theory, it is welcome from Field's standpoint. On the other hand, however, such an open-ended perspective does not correspond to limitless possibilities to recover more and more principles: the more laws and rules are validated, the more principles are also ruled out, due to paradoxes. For example, there are forms of conditional introduction which are just too strong to be consistent with Field's theory. Since this tradeoff is inevitable, perhaps one endorsing Field's theory may want to argue that the principles that are recovered are more important than those ruled out.

Moreover, it is far from clear that the logical and truth-theoretic principles validated by Field's theory constitute, just by themselves, a philosophically interesting story about the semantics of $\mathscr{L}_T^{\rightarrow}$. However, the requirement that a semantics motivated by FIELD'S PROGRAM should also be a philosophically interesting story about the language $\mathscr{L}_T^{\rightarrow}$, perhaps addressing my criteria **(W1)**-**(W3)**, should be motivated. In order to do so, and to test Field's revision theory against such requirements, we should look at Field's construction. As I have done with Kripke's theory, I will begin with an informal sketch of the construction.

3.2 Field's revision theory: informal presentation

Consider the least Kripke fixed point $\mathfrak{I}_\Phi$.[5] Every conditional $\varphi \rightarrow \psi$ is in its C-gap, and the same holds for every sentence resulting from $\varphi \rightarrow \psi$ substituting occurrences of a subformula χ of $\varphi \rightarrow \psi$ with $T\ulcorner\chi\urcorner$, or *vice versa*, an arbitrary number of times. $\mathfrak{I}_\Phi$, then, does not give us much information about conditionals.

[5]Recall that, from this chapter onwards, we are working with the language $\mathscr{L}_T^{\rightarrow}$ (unless otherwise specified), and so every reference to Kripke's theory should be understood as a reference to Kripke's theory for the language $\mathscr{L}_T^{\rightarrow}$.

However, we can consider the Kripke C-gappy evaluation generated by $\mathfrak{I}_\Phi$.[6] This function assigns to every $\mathscr{L}_T^{\rightarrow}$-sentence exactly one value between 1, 0, and 1/2. These semantic values are ordered in the obvious numerical way: $0 < 1/2 < 1$. So, given a sentence $\varphi \rightarrow \psi$, we can check whether the value of its antecedent φ is lower than or equal to the value of its consequent ψ, or whether the value of φ is greater than the value of ψ, according to the above ordering, in the evaluation generated by $\mathfrak{I}_\Phi$. Let's say that a conditional has *value* 1 *at the first revision stage* in the first case, and *value* 0 *at the first revision stage* in the second case. As Field puts it, "by and large, this new valuation for conditionals is an improvement on the old",[7] since the old one was just due to lack of treatment of *every* conditional, and so we can reasonably take it as our first revision step towards a better understanding of the semantic values of conditionals.

To preserve Kripke's treatment of non-conditional sentences, including sentences of the form $T\ulcorner\varphi\urcorner$, we do not need revision: we just take the Kripke fixed point that results from the pair $\langle P_1, Q_1\rangle$, where P_1 is the set of conditionals that are given value 1 at the first revision stage, and Q_1 is the set of conditionals that get value 0 at the first revision stage. In short, we build the Kripke fixed point that is generated from the results of the first revision stage.

But now we can reiterate the procedure that gave us the first revision stage, and perform a second revision over the fixed point $\mathfrak{I}_\Phi(P_1, Q_1)$. More precisely, we consider the Kripke C-gappy evaluation generated by $\mathfrak{I}_\Phi(P_1, Q_1)$, and we give *value* 1 *at the second revision stage* to every conditional $\varphi \rightarrow \psi$ such that φ has a value that is less than or equal to the value of ψ in the ordering resulting from that evaluation, while we give *value* 0 *at the second revision stage* to every conditional $\varphi \rightarrow \psi$ such that φ has a value that is greater than the value of ψ in the ordering resulting from that evaluation. Then, obviously, we go on as we did after the first

[6] See Definition 2.5.4.

[7] [Field 2008], p. 249.

revision step, and then again and again, *ad infinitum*. Of course, this is the outline of the process for successor revision stages, and it does not applies to limit stages, for which a specific rule is needed. Nevertheless, this quick sketch should suffice to give an idea of the general functioning of Field's revision construction.

As the revision proceeds, arguably the evaluations of conditionals become more and more accurate. The reason is that the more the revision procedure is carried forward, the more the structure of conditional sentences is "dug out" and taken into account. For example, at the first stage of revision, a conditional of the form $(\psi_1 \to \psi_2) \to (\chi_1 \to \chi_2)$ is evaluated using the values associated with $\psi_1 \to \psi_2$ and $\chi_1 \to \chi_2$ at $\mathfrak{I}_\Phi$, which is not very interesting because these two sentences are in its C-gap just because they are conditionals. At the second revision stage, however, the value of $(\psi_1 \to \psi_2) \to (\chi_1 \to \chi_2)$ depends on the relations between the semantic values of $\psi_1 \to \psi_2$ and of $\chi_1 \to \chi_2$ in the Kripke fixed point *built over the first revision stage*, and the resulting value assignment may not be uninterestingly trivial (as it was at the first stage, where every conditional is C-gappy by default), for example in the case in which ψ_1, ψ_2, χ_1, and χ_2 are sentences of $\mathscr{L}_T$. If some of ψ_1, ψ_2, χ_1, and χ_2 involve a conditional, this further substructure will only be revealed at later revision stages. Obviously, the presence of sentences such as Curry's sentence entails that there is no revision stage where every conditional is decomposed into sentences belonging only to $\mathscr{L}_T$.

I will now turn to a more precise presentation of Field's revision procedure.

3.3 Field's revision theory: formal presentation

Let's start with the main components of Field's revision construction. Recall that, for a consistent Kripke fixed point $\mathfrak{I}_\Phi(P,Q)$, the symbol $\mathscr{K}^+_{P,Q}$ denotes the Kripke C-gappy evaluation function generated by it.[8]

[8]See Definition 2.5.4.

Definition 3.3.1. (Field revision sequence)

Field revision construction is given by the function $r : Ord \times SENT_{\mathscr{L}_T^{\rightarrow}} \longmapsto \{1, 0, 1/2\}$, defined as follows.

- $$r(0,\varphi) := \begin{cases} 1, \text{ if } \varphi \text{ is } \psi \rightarrow \chi \text{ and } \mathscr{K}_{\emptyset,\emptyset}(\psi) \leq \mathscr{K}_{\emptyset,\emptyset}(\chi) \\ 0, \text{ if } \varphi \text{ is } \psi \rightarrow \chi \text{ and } \mathscr{K}_{\emptyset,\emptyset}(\psi) > \mathscr{K}_{\emptyset,\emptyset}(\chi) \\ \mathscr{K}_0(\varphi), \text{ if } \varphi \text{ is not a conditional} \end{cases}$$

  where $\mathscr{K}_0 := \mathscr{K}_{F_E^0, F_A^0}$, and $\langle F_E^0, F_A^0 \rangle$ is

  $$\mathfrak{I}_\Phi(\{\psi \rightarrow \chi \in \mathscr{L}_T^{\rightarrow} \,|\, r(0, \psi \rightarrow \chi) = 1\}, \{\psi \rightarrow \chi \in \mathscr{L}_T^{\rightarrow} \,|\, r(0, \psi \rightarrow \chi) = 0\}).$$

  Strictly speaking, two steps are conflated in the above graph bracket: *first* $\mathscr{K}_{\emptyset,\emptyset}$ evaluates conditionals, *then* $\mathscr{K}_0$ is built, being generated by the above Kripke fixed point, which uses the conditionals just evaluated using $\mathscr{K}_{\emptyset,\emptyset}$. I write together these two steps for simplicity.

- $$r(\alpha+1,\varphi) := \begin{cases} 1, \text{ if } \varphi \text{ is } \psi \rightarrow \chi \text{ and } \mathscr{K}_\alpha(\psi) \leq \mathscr{K}(\chi) \\ 0, \text{ if } \varphi \text{ is } \psi \rightarrow \chi \text{ and } \mathscr{K}_\alpha(\psi) > \mathscr{K}(\chi) \\ \mathscr{K}_{\alpha+1}(\varphi), \text{ if } \varphi \text{ is not a conditional} \end{cases}$$

  where $\mathscr{K}_{\alpha+1} := \mathscr{K}_{F_E^{\alpha+1}, F_A^{\alpha+1}}$, and $\langle F_E^{\alpha+1}, F_A^{\alpha+1} \rangle$ is

  $$\begin{aligned} &\mathfrak{I}_\Phi(\{\psi \rightarrow \chi \in \mathscr{L}_T^{\rightarrow} \,|\, r(\psi \rightarrow \chi, \alpha+1) = 1\}, \\ &\quad \{\psi \rightarrow \chi \in \mathscr{L}_T^{\rightarrow} \,|\, r(\psi \rightarrow \chi, \alpha+1) = 0\}), \end{aligned}$$

  and similarly for $\mathscr{K}_\alpha$.

$$
\bullet\quad r(\delta,\varphi) := \begin{cases} 1, \text{ if } \varphi \text{ is } \psi \to \chi \text{ and there is a } \gamma < \delta \text{ such that} \\ \quad \text{for all } \beta \text{ such that } \beta \geq \gamma \text{ and } \beta < \delta, \mathscr{K}_\beta(\psi) \leq \mathscr{K}_\beta(\chi) \\ 0, \text{ if } \varphi \text{ is } \psi \to \chi \text{ and there is a } \gamma < \delta \text{ such that} \\ \quad \text{for all } \beta \text{ such that } \beta \geq \gamma \text{ and } \beta < \delta, \mathscr{K}_\beta(\psi) > \mathscr{K}_\beta(\chi) \\ 1/2, \text{ if } \varphi \text{ is } \psi \to \chi \text{ and none of the two above cases is given} \\ \mathscr{K}_\delta(\varphi), \text{ if } \varphi \text{ is not a conditional} \end{cases}
$$

for δ a limit ordinal, where $\mathscr{K}_\beta$ is defined as above for every $\beta \leq \delta$. ▲

There is exactly one function $\mathscr{K}_\alpha$ per ordinal. In every bullet point above, the conditions on the right-hand side of the curly brackets are exhaustive and mutually exclusive. The values taken by r will be called *revision values*.

Remark 3.3.2.

Definition 3.3.1 simplifies somewhat Field's construction, eliminating some subtleties that are not relevant to get the big picture. For example, in [Field 2008], the construction of Kripke fixed points and of the Kripke C-gappy evaluations are relativized to a variable assignment, whereas I only consider closed formulae.

Moreover, it is not necessary to take $\mathscr{K}_{\emptyset,\emptyset}$ as starting point for the revision construction. One could run the revision construction over any function that Field calls *transparent evaluations for conditionals*,[9] namely those functions $\mathscr{E} : \{\varphi \in \mathscr{L}_T^{\to} \mid \varphi \text{ is a conditional}\} \longmapsto \{1, 0, 1/2\}$ such that for every conditional $\varphi \to \psi \in \mathscr{L}_T^{\to}$ and every $\varphi' \to \psi'$ resulting from $\varphi \to \psi$ substituting occurrences of some subformula χ of $\varphi \to \psi$ with $T\ulcorner\chi\urcorner$, or *vice versa*, an arbitrary number of times, we have that $\mathscr{E}(\varphi \to \psi) = \mathscr{E}(\varphi' \to \psi')$.

It is equally unnecessary to take the "small" fixed points of the form $\langle F_E^\alpha, F_A^\alpha \rangle$

[9] See [Field 2008], p. 243.

at revision stages: one could take fixed points that properly include the former (at each revision stage), again without altering the main properties of the revision.

Other variants change the evaluation clauses for conditionals: different values are assigned, but the main properties of the construction remain unaltered.[10] ▲

The revision construction is used to define a first evaluation for $\mathscr{L}_T^{\rightarrow}$-sentences, which Field calls the *ultimate evaluation*, that uses a natural tripartition of the revision process. I will denote with 1^u, 0^u, and $1/2^u$ the values of the ultimate evaluation, that are called *ultimate values*. They are not different from the numbers 1, 0, and $1/2$ used as revision values, and the superscript u is just a convention to distinguish the uses of these numbers as revision values or as ultimate values.

Definition 3.3.3. (Ultimate evaluation function, ultimate values)
Field's ultimate evaluation is the function $\mathscr{U} : SENT_{\mathscr{L}_T^{\rightarrow}} \longmapsto \{1^u, 0^u, 1/2^u\}$ defined as follows:

1. $\mathscr{U}(\varphi) = 1^u$, if there is an ordinal α such that for all $\beta \geq \alpha$, $r(\beta, \varphi) = 1$.

2. $\mathscr{U}(\varphi) = 0^u$, if there is an ordinal α such that for all $\beta \geq \alpha$, $r(\beta, \varphi) = 0$.

3. $\mathscr{U}(\varphi) = 1/2^u$, if neither of the two cases above obtains.

The result of $\mathscr{U}(\varphi)$ will be called the *ultimate value* of φ. ▲

If the ultimate value of a sentence is 1^u (0^u), we will say that its revision sequence *stabilizes on* 1 (*stabilizes on* 0); if, instead, the ultimate value of a sentence φ is $1/2^u$, we will say that its revision sequence *does not stabilize*.

The main result about ultimate values and the ultimate evaluation is the following theorem, that tells us that Field's ultimate evaluation is well-defined.

[10] One such variant will be discussed in connection to the treatment of κ in Subsection 3.5.1.

Theorem 3.3.4. (Field's fundamental theorem on ultimate values)[11]

For every ordinal α, there is a $\delta > \alpha$ such that for every $\varphi \in \mathscr{L}_T^{\rightarrow}$, $r(\delta, \varphi) = \mathscr{U}(\varphi)$.

The ordinals that respect the above theorem are called *acceptable*: they can be thought as being large enough to represent the revision process and its outcomes with suitable reliability. Theorem 3.3.4 guarantees the existence of exactly one function with the properties ascribed to $\mathscr{U}$.

Field's ultimate evaluation has various nice properties, some of which will be reviewed in the next section, but is interestingly developed into another, much more expressive evaluation for $\mathscr{L}_T^{\rightarrow}$-sentences.

The main motivation to modify the ultimate evaluation is the unsatisfactory treatment of the value $1/2^u$. As Field highlights, it "hid[es] great many distinctions", as there are crucial differences between some revision sequences that do not stabilize. For example, if the revision sequence of ψ keeps oscillating between 0 and $1/2$, and the revision sequence of χ keeps oscillating between $1/2$ and 1, the ultimate value of $\psi \rightarrow \chi$ is 1^u and the ultimate value of $\chi \rightarrow \psi$ is $1/2^u$.[12] However, the ultimate value of ψ is $1/2^u$ just as the ultimate value of χ. A conditional whose antecedent and consequent both have ultimate value $1/2^u$, then, may have ultimate value 1^u or ultimate value $1/2^u$ (but not 0^u). Similar oddities affect the relations between ultimate values 1^u and $1/2^u$, or between ultimate values $1/2^u$ and 0^u. Such problems are caused by the wild variety of ways in which a sequence can oscillate, in the very long run that it takes to reach an acceptable ordinal.

The fine-grained evaluation makes explicit this wild variety, in the most literal

[11] A proof of this result is sketched in [Field 2008], pp. 257-258.

[12] The reason for $\mathscr{U}(\psi \rightarrow \chi) = 1^u$ is that there is an ordinal α such that, for all $\beta \geq \alpha$, the revision value of ψ at β is less than or equal to the revision value of χ. The reason for $\mathscr{U}(\chi \rightarrow \psi) = 1/2^u$, instead, is that it is not the case that there is an ordinal γ such that, for all $\delta \geq \gamma$, the revision value of χ is less than or equal to the revision value of ψ at δ. Moreover, since both ψ and χ fail to stabilize, they get revision value $1/2$ at every limit stage, and thus it is also not the case that there is a γ such that, for all $\delta \geq \gamma$, the revision value of χ is strictly greater than the revision value of ψ at δ.

way. The idea behind the fine-grained version is easily described informally: instead of assigning to a sentence a semantic value that "represents" (with excessive simplification) its revision sequence, we assign to a sentence *its entire revision sequence as semantic value*. To simplify things, we will only consider the revision sequence which is generated *after* the first acceptable ordinal. Some such sequences are easy to describe. For example, the revision process of a sentence φ stabilizes on 1 if and only if φ has only revision value 1 after the first acceptable ordinal, and this happens if and only if φ has only revision value 1 after every acceptable ordinal, so the fine-grained value of φ is the function constant on revision value 1; the same happens for revision processes stabilizing on 0. Non-stabilizing sequences, instead, produce genuine oscillations for the length of the fragment of ordinals for which they are defined (such fragment should be large enough to allow the possible cycles of oscillations between revision values generated by $\mathscr{L}_T^{\rightarrow}$-sentences to be suitably represented, without being "cut").

To make this picture formally precise, the first thing to do is to specify the value space that the fine-grained evaluation employs.

Definition 3.3.5. (Field's value space)

To describe Field's fine-grained value space, consider the following ordinals:

- Let δ_0 be the least acceptable ordinal.
- Let γ_0 be the ordinal s.t. $\delta_0 + \gamma_0$ is the least acceptable ordinal after δ_0.[13]
- Let γ^+ be the smallest initial ordinal whose cardinality is greater than that of γ_0. Note that $\delta_0 + \gamma^+ = \gamma^+$.
- Let $\mathcal{Sm}(\gamma^+)$ be the set of ordinals smaller than γ^+.

Field's value space is the set $\mathbf{F}$ of all the functions $\mathfrak{f} : \mathcal{Sm}(\gamma^+) \longmapsto \{1, 0, 1/2\}$ that satisfy the following conditions:

[13] Of course, with "+" I indicate the operation of ordinal sum here.

1. If $\mathfrak{f}(0) = 1$, then for all ordinals α, $\mathfrak{f}(\alpha) = 1$.

2. If $\mathfrak{f}(0) = 0$, then for all ordinals α, $\mathfrak{f}(\alpha) = 0$.[14]

3. If $\mathfrak{f}(0) = 1/2$, then there is an ordinal $\gamma < \gamma^+$ such that for all ordinals α and β, if $\gamma \cdot \alpha + \beta < \gamma^+$, then $\mathfrak{f}(\gamma \cdot \alpha + \beta) = \mathfrak{f}(\beta)$ (γ-*cyclicity*).

Every $\mathfrak{f} \in \mathbf{F}$ is a *transfinite sequence*.[15] The (γ-*cyclicity*) condition ensures that, if a sequence never stabilizes on revision value 1 or on revision value 0, then there are arbitrarily large ordinals (within $\mathcal{S}m(\gamma^+)$) at which it takes all the values featured in its oscillation. ▲

Some noteworthy elements of Field's value space are: the function constant on revision value 1, denoted with $\mathbf{1}$ (which is the designated value), and the functions $\mathbf{0}$ and $\mathbf{1/2}$, defined similarly with revision values 0 and 1/2 respectively.

The algebraic counterpart of Field's conditional is given by the following partial ordering $\preceq$ on $\mathbf{F}$, defined pointwise: for every $\mathfrak{f}, \mathfrak{g} \in \mathbf{F}$,

$$\mathfrak{f} \preceq \mathfrak{g} :\Leftrightarrow \text{ for every ordinal } \alpha < \gamma^+, \mathfrak{f}(\alpha) \leq \mathfrak{g}(\alpha). \tag{3.1}$$

The "$\leq$" above is the usual numerical ordering, $\prec$, $\succeq$, $\succ$ are defined as usual.

Remark 3.3.6.

Also here, for the sake of simplicity, I am giving one particular version of the value space for the fine-grained semantics, but many variants of it could be employed. For example, smaller ordinals than γ^+ could be taken ([Field 2008], p. 260), but this and other tweaks do not change the main features of the semantics. Moreover,

[14]Conditions 1 and 2 may appear strange at first, but one has to remember that we are using transfinite sequences to match the fragments of Field's revision sequences that start from the first acceptable ordinal. If a sentence stabilizes on revision value 1 at the first acceptable ordinal, its revision sequence remains stably on 1 after that. A transfinite sequence matching this behavior, thus, has to take value 1 at every ordinal if it takes value 1 at the very beginning.

[15]For this notion and the relative notation, see [Jech 2002], p. 21.

the version I am presenting is the one ultimately adopted by Field – so one is allowed to presume that it is the one he found more satisfactory. ▲

Once we have described the space of all the possible revision sequences of length γ^+ after the first admissible ordinal, we can define the evaluation that employs this space as value space.

Definition 3.3.7. (Field's fine-grained evaluation)
Field's fine-grained evaluation is the function $\mathscr{F} : SENT_{\mathscr{L}_T^{\rightarrow}} \longmapsto \mathbf{F}$ defined as:

$$\mathscr{F}(\varphi) := \langle r(\delta_0 + \alpha, \varphi) : \alpha < \gamma^+ \rangle.$$

▲

Observe that the function $\mathscr{F}$ is well-defined: for every $\varphi \in \mathscr{L}_T^{\rightarrow}$, there is exactly one function $\mathfrak{f} \in \mathbf{F}$ having the characteristics assigned to $\mathscr{F}(\varphi)$ – this is a consequence of Theorem 3.3.4. Obviously, the space $\mathbf{F}$ has a strictly larger cardinality than the image of $\mathscr{F}$, due to the size of γ^+.

3.4 Properties of Field's theory

Field's revision theory, as embodied by Field's ultimate evaluation $\mathscr{U}$ and by Field's fine-grained evaluation $\mathscr{F}$, exhibits the nice properties recalled in Section 3.1.2 that were targeted by Field in devising his construction.

Field's theory inherits intersubstitutivity of truth from Kripke's theory:

Theorem 3.4.1. (Intersubstitutivity of truth for $\mathscr{U}$ and $\mathscr{F}$ (Field))
For every $\varphi, \psi, \chi \in \mathscr{L}_T^{\rightarrow}$, if ψ results from χ by substituting, possibly non-uniformly, φ with $T\ulcorner\varphi\urcorner$, or vice versa *$T\ulcorner\varphi\urcorner$ with φ, then:*

$$\mathscr{U}(\psi) = \mathscr{U}(\chi) \qquad \textit{and} \qquad \mathscr{F}(\psi) = \mathscr{F}(\chi).$$

Finally, in the interpretations of $\mathscr{L}_T^{\rightarrow}$-sentences provided by $\mathscr{U}$ and $\mathscr{F}$, several logical principles are recovered, in a sense that may be made easily clear, although not entirely unproblematically. Field's revision construction can be seen as a way to expand the standard model of the truth-free language to a model of the full language $\mathscr{L}_T^{\rightarrow}$, and the evaluation functions $\mathscr{U}$ and $\mathscr{F}$ as two definitions of the notion "having semantic value **v**" in such model – for the respective kinds of semantic values. The "truth-free language" is $\mathscr{L}_T^{\rightarrow}$ minus the predicate T, and it can be seen as just $\mathscr{L}$, since $\rightarrow$ and $\supset$ are given the same interpretation by $\mathscr{U}$ and $\mathscr{F}$ on sentences not featuring T. Field always uses a single designated value: the semantic value 1^u for $\mathscr{U}$, and the semantic value $\mathbf{1}$ for $\mathscr{F}$. The choice of the evaluation function, however, is not relevant for the sentences getting the designated value, since:

$$\{\varphi \in \mathscr{L}_T^{\rightarrow} \mid \mathscr{U}(\varphi) = 1^u\} = \{\varphi \in \mathscr{L}_T^{\rightarrow} \mid \mathscr{F}(\varphi) = \mathbf{1}\}. \tag{3.2}$$

Field also associates a relation of logical consequence with his construction, albeit without giving many details. I will denote such relation with $\models_F$. Field reads $\models_F$ as "*preserv[ing] designated value in all models*".[16] I interpret this to mean that $\models_F$ "preserves of the designated value *in every model built using Field's revision construction*", so I suggest the following formalization:

$\Gamma \models_F \psi$ if and only if for every function $\mathscr{U}_{\mathscr{M}}$ ($\mathscr{F}_{\mathscr{M}}$) defined as in Definition 3.3.3 (3.3.7) using Field's revision construction for a countable ω-model $\mathscr{M}$ of $\mathscr{L}$, if for every $\varphi \in \Gamma$ $\mathscr{U}_{\mathscr{M}}(\varphi) = 1^u$ (or $\mathscr{F}_{\mathscr{M}}(\varphi) = \mathbf{1}$), then also $\mathscr{U}_{\mathscr{M}}(\psi) = 1^u$ (or $\mathscr{F}_{\mathscr{M}}(\psi) = \mathbf{1}$).

[16] [Field 2008], p. 267.

There is exactly one countable ω-*model* of first-order arithmetic (up to isomorphism), the standard model, $\mathbb{N}$. So if the definition of $\models_F$ quantifies over more than one evaluation, it quantifies over the functions $\mathscr{U}_{\mathscr{M}}$ or $\mathscr{F}_{\mathscr{M}}$ that are just like $\mathscr{U}$ or $\mathscr{F}$, with the possible exception of using a different Kripke fixed point at revision stages. Field is never explicit as to which Kripke fixed points are usable for his purposes,[17] and it is natural to suppose that some Kripke fixed point assign values to conditional sentences in a really bad way, producing unwanted revision sequences. However, we can suppose that the Fieldian theorist is able to put a restriction on the quantification over functions $\mathscr{U}_{\mathscr{M}}$ or $\mathscr{F}_{\mathscr{M}}$, if there need be. We can use $\models_F$ to state that several logical laws and rules are recovered in Field's theory.

Theorem 3.4.2. (Some important principles recovered by $\mathscr{U}$ and $\mathscr{F}$)

- $\models_F \varphi \to \varphi$
- $\models_F \neg\neg\varphi \leftrightarrow \varphi$
- $\models_F \varphi \wedge \psi \to \varphi$; $\models_F \varphi \wedge \psi \to \psi$
- $\models_F \varphi \to \varphi \vee \psi$; $\models_F \psi \to \varphi \vee \psi$
- $\models_F (\varphi \to \psi) \leftrightarrow (\neg\psi \to \neg\varphi)$
- $\models_F \varphi \wedge (\psi \vee \chi) \to (\varphi \wedge \psi) \vee (\varphi \wedge \chi)$
- $\models_F \neg(\varphi \wedge \psi) \leftrightarrow \neg\varphi \vee \neg\psi$
- $\models_F \neg(\varphi \vee \psi) \leftrightarrow \neg\varphi \wedge \neg\psi$
- $\models_F \forall x\varphi(x) \to \varphi(x/t)$
- $\models_F \varphi(x/t) \to \exists x\varphi(x)$
- $\varphi, \psi \models_F \varphi \wedge \psi$
- $\varphi, \varphi \to \psi \models_F \psi$
- $\varphi, \neg\psi \models_F \neg(\varphi \to \psi)$
- $\varphi \models_F \psi \to \varphi$
- $\varphi \to \psi \models_F (\psi \to \chi) \to (\varphi \to \chi)$
- $(\varphi \to \psi) \models_F (\varphi \wedge \chi) \to (\psi \wedge \chi)$
- $(\varphi \to \chi) \wedge (\psi \to \chi) \models_F \varphi \vee \psi \to \chi$
- $(\varphi \to \psi) \models_F (\varphi \vee \chi) \to (\psi \vee \chi)$
- $(\varphi \to \psi) \wedge (\varphi \to \chi) \models_F \varphi \to \psi \wedge \chi$
- $\varphi(x) \models_F \forall x\varphi(x)$.

[17] See [Field 2008], p. 249.

This result completes the essentials of Field's revision theory: let's turn now to some critical considerations on it.

Field's theory has raised quite a lot of reactions in the truth-theoretic literature. Now, it is true that Field conceives his model primarily as an instrument to show that several logical principles (that he considers desirable) are consistent with intersubstitutivity of truth. Nevertheless, results such as Theorems 3.4.1 and 3.4.2 show that a theory of naïve truth can be very strong and expressive, both from truth-theoretic and a logical standpoint. In addition, Field's theory suggests sophisticated solutions to important philosophical problems, including the treatment of truth-theoretic paradoxes.[18] So, it is natural to look for a more general understanding of the picture provided by the semantics, and for its conceptual consequences as well. Analyzing Field's theory, however, several authors complained that the construction is quite complicated and somewhat unnatural.[19] For example, here is how Beall puts his reasons for dissatisfaction:

> [...] I do not know of any terribly strong arguments against Field's approach. My main reason for preferring my own account comes down, I'm afraid, to a fairly fuzzy sense that Field's approach [...] might [...] be more complicated than we need. Regrettably, I do not know how to make the relevant sense of complexity [...] precise enough to serve as an objection (it certainly isn't merely a matter of standard mathematical accounts of complexity). ([Beall 2009], Preface, p. viii.)

In what follows, I will try to make the (sometimes vague) puzzlement and worries expressed by many philosophers and logicians into some specific critical remarks on Field's theory, as a philosophical story about the semantics of $\mathscr{L}_T^{\rightarrow}$, as a means to establish the consistency of some principles, and as an ultimately satisfactory solution to truth-theoretic paradoxes and other crucial problems.

[18] The philosophically crucial issues addressed by Field's revision theory include the notion of *determinate* truth. I discuss this point in Section 3.6.

[19] For a debate on this, see for example [D.A. Martin 2011], [Welch 2011] and [Field 2011].

A good place to start a critical investigation of Field's theory is the question of *what is Field's conditional*.

3.5 What is Field's conditional?

In order to motivate the question that gives the title to this section, let us consider Field's theory as it is meant to be, namely as a way to show that some desirable logical principles are consistent with intersubstitutivity of truth. Another question arises immediately: *why are they desirable*?

It is quite natural to ask the latter question in the context of naïve truth. In fact, if one claims some *logical* principles to be desirable (and not others), she should explain why this is so, otherwise we may have no reason to want them. Due to naïve truth, in fact, we have to abandon several long-standing logics (classical, intuitionistic, and so on) with their well-established conceptual foundations: this makes particularly urgent the need to justify the rather special and limited set of principles validated by theories of naïve truth.

The situation is somewhat different for *truth-theoretic* principles. Truth theorists may agree on some formulations of naïveté, while disagreeing fiercely about the logical principles naïve truth should go with. Moreover, a naïve notion of truth may need no particular justification in itself, since it draws much of its appeal from the ordinary use of the word "true" – or so many authors think. Hardly anything analogous to the latter point can be said for conditional principles. In fact, the appeal to natural language does not seem to play the same role in motivating naïve truth and in motivating conditional principles, given the well-known divergences that arise between conditional statements in mathematics and conditional statements in natural languages *tout court*.[20]

[20]Considerations specific to the combination of naïve truth and logical principles are a much more cogent guide of what can be accepted or should be rejected in theories of truth. For example, [Field 2008] (p. 268) rightly stresses that, if one accepts naïveté, the usual structural rules, and *modus*

So, the question of why Field's principles are desirable seems well-motivated. How can we answer it? A simple answer is that Field's principles look very *intuitive*, in the present context, since they are classically valid. However, it is obvious that, in a statement of the form $\varphi \rightarrow \varphi$, any specific φ does not play any role in our acceptance of the principle. The reason why we accept it, if we do, must come from the conditional, the only invariant element of the law (similar remarks go for the other schemata). If $\varphi \rightarrow \varphi$ looks intuitive, there must be something in the understanding of the $\rightarrow$ occurring in the law that makes the statement $\varphi \rightarrow \varphi$ to come out pretty natural. The Fieldian theorist, thus, should give an account of the conditional that makes it clear why $\varphi \rightarrow \varphi$ comes out as intuitive as it does. Therefore, also the more general question of what is Field's conditional is straightforwardly motivated by reflecting on Field's theory and the logical principles it validates.

An account of connectives and quantifiers is usually provided by the semantics itself. Field himself concedes the need to provide an understanding of the connectives and quantifiers, via the model-theoretic semantics, and he contrasts the case of connectives and quantifiers with the case of truth.

> In my view, model theory plays at best a very indirect role in explaining truth. Rather, truth is directly explained by means of Schema (T), and *model theory enters in only in helping us understand the logical connectives* that occur in instances of Schema (T). [...] ([Field 2007], p. 102, my emphasis)

So, it is natural to require that the model-theoretic construction should provide a reading for the conditional that justifies the acceptance of the principles that hold in the models developed using the construction.

Unfortunately, this is difficult for Field's theory: several common ways of explaining what a conditional connective is are blocked, and the model does not seem to provide a new one, as the following analysis shows (Subsections 3.5.1-3.5.3).

ponens, Curry's paradox forces us to reject the rule of conditional contraction ("from $\varphi \rightarrow (\varphi \rightarrow \psi)$, infer $\varphi \rightarrow \psi$"), and the corresponding law ($(\varphi \rightarrow (\varphi \rightarrow \psi)) \rightarrow (\varphi \rightarrow \psi)$").

3.5.1 Comparing semantic values?

In the context of many-valued logics, an interesting possibility is to see the conditional as a *tool to compare semantic values*. In this interpretation, a conditional is associated with an ordering of semantic values, and it is supposed to have the designated value if the value of its antecedent is less than or equal to the value of its consequent (in that ordering), and a value that expresses the difference in semantic values between its antecedent and consequent (in that ordering) otherwise. As there can be several ways to "express the difference in semantic values", I do not specify this interpretation any further, but this characterization of "a tool to compare semantic values" seems intelligible nonetheless.[21]

Reading the conditional as a tool to compare semantic values is very much in line with a quite standard way of seeing the other quantifiers and connectives of $\mathscr{L}_T^{\rightarrow}$ in many-valued semantics, namely as providing information about order-theoretic relations concerning the semantic values of the immediate subsentences of negations, conjunctions, disjunctions, and universally quantified sentences.

For example, in several readings, the semantic value of a disjunction provides information about the relations that hold between the semantic values of the two disjuncts and semantic values that are *greater than or equal to* them, in terms of some partial ordering of the semantic values. More specifically, in a setting where semantic values are linearly ordered (and there is only one designated value), in several semantics, a disjunction takes the value (amongst the semantic values of the disjuncts) that is *closer to* the designated one – with the *proviso* that, if the two disjuncts have the same value, they are equally close to the designated value, and the disjunction gets the same semantic value as the disjuncts.[22] Alternatively, if the semantic values are only partially ordered, in several readings, the semantic value

[21] A fundamental source on algebraic approaches to non-classical logics is [Rasiowa 1974].

[22] This is the case of Kripke's theory, for example. Depending on the specific setting, this reading can accommodate also multiple designated semantic values, but I will not consider this case here.

of a disjunction is understood as the least upper bound (if it exists) of the semantic values of the disjuncts in that ordering.[23]

Kripke's and Field's theories adopt order-theoretic readings of the kind just described for $\neg$, $\wedge$, $\vee$, and $\forall$.[24] So, I take it, order-theoretic readings of connectives and quantifiers are not only common in logic, but also quite relevant in the setting of naïve truth. Interpreting a conditional as a tool to compare semantic values in the above sense, then, would contribute significantly to understanding and expressing the semantic value relations between the sentences of $\mathscr{L}_T^{\rightarrow}$ that also the other connectives and quantifiers aim at conveying.

Now, in defining the values of a conditional at revision stages, Field employs a comparison of revision values, but it is impossible to understand the conditional as comparing final values. We have seen that ultimate values will not do for this purpose, as there are sentences ψ_1, ψ_2, χ_1, χ_2 such that they all have ultimate value $1/2^u$ but $\psi_1 \rightarrow \chi_1$ has ultimate value 1^u while $\psi_2 \rightarrow \chi_2$ has ultimate value $1/2^u$.[25] Thus, to decide whether Field's $\rightarrow$ performs a comparison of values, we must turn to the fine-grained version of Field's construction.

In this semantics, as we have seen, the conditional is algebraically interpreted by the partial order $\preceq$, defined by statement (3.1). [Leitgeb 2007] shows that, even if a sentence $\varphi \rightarrow \psi$ gets value $\mathbf{1}$ in Field's theory if and only if the value of φ is greater than or equal to the value of ψ in Field's partial order $\preceq$, it is not the case that $\varphi \rightarrow \psi$ receives value $\mathbf{0}$ if and only if the value of φ is not greater than nor equal to the value of ψ, in the same partial order. However, this sense of value comparison is too much to ask if we have more than 2 semantic values: as Leitgeb notices, in the Curry case the value of Tt_κ is not greater than nor equal to the value

[23] The truth-ordering in the four-valued lattice for First Degree Entailment is an example of such a case (see [Belnap 1977], pp. 13-16).

[24] For Kripke's theory this is quite obvious; for Field's theory, see the beginning of Section 3.6, and [Field 2008], pp. 259-260

[25] A similar problem affects the theory in [Field 2014].

of $0 \neq 0$ (in $\preceq$) but it would be disastrous if the whole sentence had value **0**.

Nevertheless, having only conditions to determine when a conditional has semantic value **1** is not satisfactory. In order to interpret Field's conditional as a tool to compare semantic values, we must be able to say something about the semantic value of a sentence of the form $\varphi \rightarrow \psi$ in terms of the relation between the values of φ and ψ also when the whole conditional does not have semantic value **1**. To see whether this is possible we must look into Field's value space **F**.[26]

As recalled after Definition 3.3.5, the value space **F** has some noteworthy elements, namely the function constant on 1, in symbols **1**, and the functions **0** and **1/2**, defined similarly with 0 and 1/2. In addition to these values, Field identifies some semantic value "zones" within **F**: let's say that $\mathscr{F}(\varphi) \in \mathbf{H}$ (**L**, **E**) if the revision values of φ oscillate between 1/2 and 1 (0 and 1/2; 1, 1/2 and 0 respectively). We can read **H** as the "higher" part of **F**, **L** as the "lower" one, while **E** is "external" to such relations – the choice of these names will be justified shortly.

All the possible $\rightarrow$-combinations of sentences receiving the semantic values just described yield the following table:

$\rightarrow$	**0**	**L**	**1/2**	**H**	**1**	**E**
0	**1**	**1**	**1**	**1**	**1**	**1**
L	**E**	**E; 1**	**1**	**1**	**1**	**E; 1**
1/2	**0**	**E**	**1**	**1**	**1**	**E**
H	**0**	**E**	**E**	**E; 1**	**1**	**E**
1	**0**	**0**	**0**	**E**	**1**	**E**
E	**E**	**E**	**E**	**E; 1**	**1**	**E; 1**

Table 3.1: Semantic values and zones comparison table for Field's $\rightarrow$

"$\mathbf{E}; \mathbf{1}$" means that either $\mathscr{F}(\varphi \rightarrow \psi) \in \mathbf{E}$ or $\mathscr{F}(\varphi \rightarrow \psi) = \mathbf{1}$. The table shows

[26]See Definition 3.3.5.

that $\mathbf{H},\mathbf{L},\mathbf{E}$ hide a great variety of sequences, that are $\rightarrow$-combined with other sequences in many different ways.

In particular, zones **H** and **L** are "between" values **0** and **1**:

(3.3) $$\mathbf{0} \prec \text{ any value in } \mathbf{L} \prec \mathbf{1/2} \prec \text{ any value in } \mathbf{H} \prec \mathbf{1}.$$

As Field notices, moreover, values in **E** do not fit in the relation (3.3), and there is much less that we can say about them in terms of the ordering $\preceq$:

(3.4) $$\mathbf{0} \prec \text{ any value in } \mathbf{E} \prec \mathbf{1}.$$

Unfortunately, relations (3.3) and (3.4) exhaust what can be said about comparisons of semantic values via $\rightarrow$ in Field's theory. The following remarks show why we cannot understand Field's $\rightarrow$ as a tool to compare semantic values.

(a) Conditionals in **E** express failure of comparability between the values of antecedent and consequent. In fact, it is easily seen that

$$\mathscr{F}(\varphi \rightarrow \psi) \in \mathbf{E} \text{ if and only if } \mathscr{F}(\varphi) \npreceq \mathscr{F}(\psi) \text{ and } \mathscr{F}(\varphi) \nsucc \mathscr{F}(\psi).$$

The importance of the value zone **E** is not to be underestimated: as Field puts it, values in **E** "play a crucial role in the theory. [...] any conditional that doesn't have value **1** or **0** must clearly have a value in [**E**]".[27]

What is crucial, however, is that there are sentences χ_1 and χ_2 such that $\mathscr{F}(\chi_1 \rightarrow \chi_2) \in \mathbf{E}$ and $\mathscr{F}(\chi_2 \rightarrow \chi_1) \in \mathbf{E}$.[28] It is highly problematic to extract

[27][Field 2008], p. 261.

[28]Note that one cannot express the fact that $\mathscr{F}(\chi_1 \rightarrow \chi_2) \in \mathbf{E}$ and $\mathscr{F}(\chi_2 \rightarrow \chi_1) \in \mathbf{E}$ by saying that $\mathscr{F}(\chi_1 \leftrightarrow \chi_2) \in \mathbf{E}$, even if the latter is a consequence of the former: in fact, it could be that $\mathscr{F}(\varphi \leftrightarrow \psi) \in \mathbf{E}$ because $\mathscr{F}(\varphi \rightarrow \psi) \in \mathbf{E}$ but $\mathscr{F}(\psi \rightarrow \varphi) = \mathbf{1}$. On the contrary, if $\mathscr{F}(\varphi \leftrightarrow \psi) = \mathbf{1}$, then $\mathscr{F}(\varphi \rightarrow \psi) = \mathbf{1}$ and $\mathscr{F}(\psi \rightarrow \varphi) = \mathbf{1}$.

information about the relationships between the values of such sentences. Field himself says that the values in $\mathbf{E}$ are "*incomparable* with $\mathbf{1/2}$" because if $\mathscr{F}(\varphi) \in \mathbf{E}$ and $\mathscr{F}(\psi) = \mathbf{1/2}$, then both $\mathscr{F}(\varphi \to \psi)$ and $\mathscr{F}(\psi \to \varphi)$ are in $\mathbf{E}$.[29] By Field's own verdict, thus, sentences such as χ_1 and χ_2 are *incomparable*.

A common way to represent an order between the values of two sentences is assigning values in a numerical domain D, considering the ordering $\leq_D$ associated with that numerical domain, and checking whether the first value is less than or equal to the second one, or whether it is greater than the second one instead. This association is impossible in Field's theory. Consider χ_1 and χ_2 as above, and suppose that we associate i with χ_1 and j with χ_2, for i and j in a numerical domain D. Then, either $i \not\leq_D j$ and $i \not>_D j$ (which seems absurd) since $\mathscr{F}(\chi_1 \to \chi_2) \neq \mathbf{1}$ and $\mathscr{F}(\chi_2 \to \chi_1) \neq \mathbf{1}$, or we accept that $\to$ doesn't express this numerical comparison, for the same reason.

Claiming that sentences in $\mathbf{E}$ have no value is not an option: Field's theory is *total*, i.e. it gives a value to every sentence of $\mathscr{L}_T^{\to}$, and this is essential to recover the schemata he wants.

On a different note, every paradoxical sentence involving the conditional is in $\mathbf{E}$, thus making it difficult to compare it with other paradoxical sentences. Take for example κ. Curry's sentence is absolutely incomparable with its negation, i.e. $\mathscr{F}(\kappa \to \neg\kappa) \in \mathbf{E}$ and $\mathscr{F}(\neg\kappa \to \kappa) \in \mathbf{E}$, in contrast to the simple treatment of the liar: $\mathscr{F}(\lambda) = \mathbf{1/2} = \mathscr{F}(\neg\lambda)$ and $\mathscr{F}(\lambda \leftrightarrow \neg\lambda) = \mathbf{1}$. This can be remedied in a variant of Field's theory in which κ has the same revision sequence as λ.[30] However, if one adopts naïveté and argues that λ and κ should be treated in the same way, then she presumably accepts that negating a sentence is equivalent to saying that from that sentence a

[29] Field [Field 2008], p. 261, my emphasis.

[30] Of course, this is the effect of a general modification of the construction, it is not an *ad hoc* alteration just for λ and κ. See [Field 2008], p. 271.

falsity follows via the conditional, as T is treated naïvely in the evaluation of both λ and κ. But this general fact does not hold in Field's modified theory either: for some φ, $\varphi \to 0 = 1$ and $\neg\varphi$ get a different value in the modified construction. Moreover, the value space of the modified construction is still **F**, so the more general considerations involving the ordering relations in **F** carry over to the modified theory as well.

(b) Clearly, some conditionals between two sentences mapped to **E** have value **1**.[31] However, not only it is not possible to express the conditions to assign value **1** or a value in **E** to such conditionals assigning numerical values (as seen), we cannot even give such conditions in terms of position within the space **F**. The reason for this is that for every $\mathfrak{e}_1 \in \mathbf{E}$ there are $\mathfrak{e}_2, \mathfrak{e}_3 \in \mathbf{E}$ such that $\mathfrak{e}_1 \preceq \mathfrak{e}_2$ but $\mathfrak{e}_1 \npreceq \mathfrak{e}_3$ and $\mathfrak{e}_1 \nsucceq \mathfrak{e}_3$. Conditionals between sentences having values such as $\mathfrak{e}_1$ and $\mathfrak{e}_2$ get value **1** in at least one direction, while conditionals between sentences having values such as $\mathfrak{e}_1$ and $\mathfrak{e}_3$ will get a value in **E** in both directions. It is easy to see that for every φ such that $\mathscr{F}(\varphi) \in \mathbf{E}$, there are sentences ψ and χ satisfying the conditions of functions $\mathfrak{e}_1$, $\mathfrak{e}_2$, and $\mathfrak{e}_3$ above. Since every function in **E** is comparable in one direction with at least one function and incomparable with another function, we cannot distinguish sub-spaces of **E** that always generate value **1** for some conditionals. The same point holds for **L** and **H**, as Table 3.1 shows.

(c) If one maintains that $\to$ performs a semantic value comparison, she should explain why value **1/2** and the zones **H** and **L** are "unreachable" by any $\to$-comparison, as no conditional gets such values. This seems strange if we consider that, thanks to relation (3.3), sentences valued **1/2** or in **H** or **L** can be compared "upwards" between them. This seems a mere technical drawback, without any conceptual support.

[31] Value **0** is excluded because of (γ-*cyclicity*), condition 3 of Definition 3.3.5.

(d) Analogously, as we have seen, sentences having value $\mathbf{1/2}$ and sentences in $\mathbf{E}$ are incomparable via $\rightarrow$. How should we understand $\mathbf{1/2}$ and $\mathbf{E}$ to account for their radical unrelatedness?

(e) Finally, even when the semantics identifies some positive information from the negative fact that $\mathscr{F}(\varphi) \npreceq \mathscr{F}(\psi)$, i.e. when $\mathscr{F}(\varphi) \succ \mathscr{F}(\psi)$, very little can be said in terms of value comparison via $\rightarrow$. Let $\mathscr{F}(\varphi) = \mathbf{1}$, $\mathscr{F}(\psi_1) = \mathbf{0}$ and $\mathscr{F}(\psi_2) \in \mathbf{L}$ or $\mathscr{F}(\psi_2) = \mathbf{1/2}$. The difference in value between φ and ψ_1 is greater than the one between φ and ψ_2 according to the ordering (3.3), but Field semantics cannot see it: $\mathscr{F}(\varphi \rightarrow \psi_1) = \mathscr{F}(\varphi \rightarrow \psi_2) = \mathbf{0}$. Also, Field's semantics cannot express such difference: consider the sentence

$$(\varphi \rightarrow \psi_2) \rightarrow (\varphi \rightarrow \psi_1). \tag{3.5}$$

If $\rightarrow$ compares semantic values, we should expect (3.5) to express that the value difference between φ and ψ_1 is strictly greater than the one between φ and ψ_2 by giving to the entire sentence a value lower than $\mathbf{1}$ in Field's ordering. However, this is not the case: $\mathscr{F}((\varphi \rightarrow \psi_2) \rightarrow (\varphi \rightarrow \psi_1)) = \mathbf{1}$. There are more limitations of this kind: by the ordering (3.3), value $\mathbf{1}$ is is strictly greater than the values in $\mathbf{H}$ (in Field's ordering $\preceq$), but the conditional cannot be used to express this fact: if $\mathscr{F}(\varphi) = \mathbf{1}$ and $\mathscr{F}(\psi) \in \mathbf{H}$, then $\mathscr{F}(\varphi \rightarrow \psi) \in \mathbf{E}$. This is a consequence of Field's construction, but it is unjustified in the light of (3.3) and of the fact that a conditional whose antecedent has value $\mathbf{1}$ and whose consequent has a value strictly less than $\mathbf{1}$ but not in $\mathbf{H}$ can at least be given a value (i.e. $\mathbf{0}$) that indicates that the value of the antecedent is greater than the value of the consequent (in $\preceq$), albeit with no gradation. This problem is not limited to $\mathbf{1}$ and $\mathbf{H}$, but presents itself for every conditional whose antecedent has a value in the ordering (3.3) and whose consequent has the value or is in the value zone immediately preceding the value of the antecedent in the ordering (3.3) (as Table 3.1 shows).

A natural reply to these difficulties would be to look for another way to order the value space $\mathbf{F}$, different from $\preceq$, but this reply is a blind alley. The previous discussion shows that Field's $\rightarrow$ does not obey the so-called "Dummett's law":

$$(\varphi \rightarrow \psi) \vee (\psi \rightarrow \varphi).$$

For $\mathscr{F}((\varphi \rightarrow \psi) \vee (\psi \rightarrow \varphi)) = \mathbf{1}$ if and only if $\mathscr{F}(\varphi \rightarrow \psi) = \mathbf{1}$ or $\mathscr{F}(\psi \rightarrow \varphi) = \mathbf{1}$, but we have seen cases where the latter disjunction fails. Therefore, any total ordering $\preceq_{tot}$ of $\mathbf{F}$ would not be an algebraic counterpart of $\rightarrow$, as we would have that, for every $\mathfrak{f}, \mathfrak{g} \in \mathbf{F}$, either $\mathfrak{f} \preceq_{tot} \mathfrak{g}$ or $\mathfrak{g} \preceq_{tot} \mathfrak{f}$.

3.5.2 Conditional introduction in Field's revision theory

An interesting aspect of the meaning of a connective is given by its *introduction and elimination rules*. Famously, some authors claim that such rules exhaust the meaning of connectives and quantifiers. In the context of the present work, of course, I mean *semantic* rules, i.e. rules given in terms of some semantic consequence relation (e.g. $\models_F$ as used in Theorem 3.4.2), since several semantic theories of truth lack an axiomatic counterpart.

Anyone who holds that introduction and elimination rules characterize connectives and quantifiers, or who attaches some importance to such rules will presumably be unhappy with Field's conditional. In fact, although *modus ponens* is validated, as $\varphi, \varphi \rightarrow \psi \models_F \psi$ holds for every $\varphi, \psi \in \mathscr{L}_T^{\rightarrow}$, conditional introduction is quite unsatisfactory. Now, in the context of naïve truth, if one has *modus ponens*, she cannot consistently keep the classical conditional introduction, since it is inconsistent with *modus ponens* and intersubstitutivity of truth, as Curry's paradox shows.[32] Still, as emphasized in the previous chapter,[33] it seems highly desirable to have a weakening of the classical conditional introduction which is conceptually interesting, to determine the relation between a conditional assertion and the assertion of a conditional, and as strong as possible.

Field observes that, in his theory, we can introduce a conditional when the excluded middle holds for the antecedent:[34]

$$\text{from } \Gamma, \varphi \models_F \psi \text{ and } \models_F \varphi \vee \neg\varphi, \text{ infer } \Gamma \models_F \varphi \rightarrow \psi. \tag{3.6}$$

But this is not very interesting. This rule is very weak: it is not even specific to Field's construction, as an identical rule holds for the introduction of $\supset$ in Kripke's models.[35] According to the conditions in (3.6), Kripke's and Field's theories can introduce conditionals in the same circumstances. But they can also eliminate a conditional in the same circumstances, as *modus ponens* for $\supset$ holds in Kripke's models.[36] So, if the introduction and elimination rules play a significant role in understanding one's conditional, Field's theory affords us no conceptual advantage over Kripke's theory, despite the fact that Field's conditional and Kripke's conditional are obviously very different.

In addition, (3.6) fails to cover some cases of conditionals that should be introduced according to Field's theory itself, e.g.:

$$\lambda \models_F \neg\lambda \text{ but } \not\models_F \lambda \vee \neg\lambda; \text{ still } \models_F \lambda \rightarrow \neg\lambda.$$

[32] See [Field 2008], pp. 281-283. Let me emphasize that I do not consider substructural notions of logical consequence, nor theories that accept inconsistencies.

[33] See Subsection 2.5.2 and Sections 2.6 and 2.7.

[34] [Field 2008], p. 269. My formulation is slightly different from Field's one, but it seems clearer. See also [Field 2003a], pp. 152-153.

[35] See rule (2.1), in Subsection 2.5.2. Also, a proof-theoretic version of this rule holds in the theory PKF formulated in natural deduction. For PKF, see [Halbach and Horsten 2006], for its natural deduction version see [Horsten 2011], p. 188.

[36] Of course, in the case of Kripke's theory, I am referring to some notion of logical consequence of the kind I introduced in Definition 2.5.7.

In the previous chapter, I stressed the importance of having a *genuine* condition for conditional introduction, in a theory of naïve truth that improve on Kripke's treatment of conditional.[37] Let me quickly recall that a genuine condition for conditional introduction is given (for a logical consequence relation $\models^*$ and a conditional $\rightarrow^*$) when the following equivalence holds:

$$\Gamma, \varphi \models^* \psi \text{ and } \Gamma \models^* \mathcal{H}(\varphi, \psi) \quad \text{if and only if} \quad \Gamma \models^* \varphi \rightarrow^* \psi;$$

where $\mathcal{H}(\varphi, \psi)$ is an $\mathscr{L}_T^{\rightarrow}$-sentence indicating a condition on φ, ψ, or both. In addition, I required that neither $\mathcal{H}(\varphi, \psi)$ nor any of its proper subschemata (possibly buried into instances of T) are equivalent to the $\models^*$-validity of all the corresponding conditionals, as this would render the condition uninterestingly trivial.

As noted in connection with Kripke's theory, a genuine condition for conditional introduction eliminates undesirable candidates. This is also the case for Field's theory; the following rule, for example, does not provide a genuine condition for conditional introduction:

$$\Gamma, \varphi \models_F \psi \text{ and } \Gamma \models_F (\varphi \vee \neg\varphi) \vee T\ulcorner \psi \vee (\varphi \rightarrow \psi)\urcorner \text{ if and only if } \Gamma \models_F \varphi \rightarrow \psi.$$

This rule is uninteresting since, amongst its conditions to introduce a conditional, it includes the fact that the entire conditional receives the designated value. More in particular, the problem with this condition is in the treatment of non-classical values. Let's ignore the truth predicate, as Field's theory validates the intersubstitutivity principle. In $(\varphi \vee \neg\varphi) \vee \psi \vee (\varphi \rightarrow \psi)$, the disjuncts $(\varphi \vee \neg\varphi)$ and ψ cover the cases of **1**-or-**0**-valued antecedents and of **1**-valued consequents, but while the disjunct $\varphi \rightarrow \psi$ should cover the remaining non-classical cases, in fact it states the desired result outright, namely that the resulting conditional has value **1**.

[37] See Definition 2.7.1 and the subsequent remarks.

As remarked in Section 2.7, a genuine condition for conditional-introduction has a remarkable conceptual relevance, since it seems central to understand the relation between conditional assertions and assertions of a conditional. However, it seems difficult to find a genuine condition for conditional-introduction for Field's theory. To my knowledge, Field never claims that a rule stronger than (3.6) is available. Moreover, the fact that we cannot identify a sub-space of always comparable semantic values within **E** (item (b) of Subsection 3.5.1) seems to doom the prospect to find a genuine condition. This has serious conceptual consequences: if no genuine condition for conditional-introduction can be found for Field's theory, the nature of the conceptually fundamental relation between conditional assertion and assertion of a conditional in Field's semantics is bound to remain obscure.

We have seen that the highly complicated value space of Field's theory forbids us from understanding Field's conditional as a tool to compare semantic values, and contributes to the obscurities that surround the relations between conditional assertion and assertion of the conditional. Now I will explore another way in which the high complexity of Field's theory makes it difficult to understand and characterize its conditional. *Contra* Beall, this will be the effect of a "fairly standard mathematical account of complexity", namely computational complexity.

3.5.3 Failure of adequate axiomatizability

Even if, in general, the value of $\varphi \rightarrow \psi$ is not the result of a comparison of the values of φ and ψ, for those sentences whose logical forms are related in certain ways, such comparison is always possible. For example, if ψ is φ, the revision sequences of the antecedent and consequent are identical, so sentences of the form $\varphi \rightarrow \varphi$ have always value $\mathbf{1}$. Similar considerations go for schemata $\varphi \wedge \psi \rightarrow \varphi$, $\varphi \rightarrow \varphi \vee \psi$, and all the laws that Field's semantics validate. This suggests that we could adopt the validity of some schemata as a key to understand Field's conditional: in other words, we could aim for an *axiomatization* of Field's construction. Some passages suggest that Field himself would also want to have an axiomatization of his semantic theory,[38] but also this path presents great difficulties.

Obvious complexity reasons show that there is no complete, recursively enumerable axiomatization of any theory including the truth-set of arithmetic. Since this is the case for virtually every semantic theory of truth, a widely adopted criterion to establish that a theory axiomatizes a semantic theory of truth is *adequacy*.

Definition 3.5.1. (Adequacy)[39]

Let Th be a recursively enumerable set of $\mathscr{L}_T^{\rightarrow}$-sentences, let $\mathscr{A}$ be a class of models for $\mathscr{L}_T^{\rightarrow}$. Th is an *adequate axiomatization* of $\mathscr{A}$ if and only if for all $S \subset \omega$ we have that:

$$(\mathbb{N}, S) \models \mathsf{Th} \quad \text{if and only if} \quad (\mathbb{N}, S) \in \mathscr{A}.$$

▲

Field's revision theory does not admit an adequate first-order axiomatization. The crucial reason for this fact is the high computational complexity of the set of sentences having value **1** in Field's theory, that was determined by [Welch 2009]. Let's call the set $\{\varphi \in \mathscr{L}_T^{\rightarrow} \mid \mathscr{F}(\varphi) = \mathbf{1}\}$ "Field's truth-set".

Theorem 3.5.2.

There is no first-order-order, recursively enumerable theory that axiomatizes adequately Field's truth-set.

Proof sketch.

[Welch 2009] proves that the computational complexity of Field's truth-set exceeds

[38] See for example [Field 2007], pp. 102-103.

[39] [Fischer, Halbach, Speck, Stern 2015] gives a detailed analysis of adequacy criteria. As they show, the criterion adopted here has a preeminent role between the proposed definitions of adequacy.

the complexity of Π^1_1-complete subsets of ω. Reasoning as in [Fischer, Halbach, Speck, Stern 2015], suppose that there is a first-order, recursively enumerable theory F such that:

$$(\mathbb{N}, P) \models \mathsf{F} \quad \text{if and only if} \quad P \text{ is Field's truth-set}$$

But then we would have a Δ^1_1 definition of Field's truth-set, which is absurd. $\square$

One could hope to capture syntactically Field's construction by accepting infinitary theories: Field's semantics, however, is too complex even for ω-logics.

Definition 3.5.3. **(ω_{L}-logic)**[40]

Let L be a recursively enumerable set of first-order principles that consist of some logical principles and the two following rules for truth:

$$\frac{\varphi}{T\ulcorner\varphi\urcorner}\ (T\text{-Intro}) \qquad\qquad \frac{T\ulcorner\varphi\urcorner}{\varphi}\ (T\text{-Elim})$$

such that PA formulated in $\mathscr{L}_T^{\rightarrow}$ with the logic L plus (T-Intro) and (T-Elim) is consistent (at least one such L exists).[41] Fix a non-repeating enumeration of all the closed terms of the language, which I indicate informally with $t_1, t_2, \ldots$. Consider the following rule:

$$\frac{\varphi(t_0) \qquad \varphi(t_1) \qquad \ldots \qquad \varphi(t_n) \qquad \ldots}{\forall n \varphi(t_n)}\ (\omega\text{-Rule})$$

An ω_{L}-*derivation* is a well-ordered sequence of $\mathscr{L}_T^{\rightarrow}$-sentences $\langle \psi_1, \ldots, \psi_\alpha, \ldots \rangle$ such that each member of the sequence ψ_β is:

- an atomic or negated atomic arithmetical sentence true in $\mathbb{N}$, or
- an instance of a logical schema of L, or

[40] See [McGee 1991], p. 150.

[41] One should want these rules, at the very least, to attempt a characterization of Field's theory.

- obtained from some elements $\langle \psi_\gamma |$ for some $\gamma < \beta \rangle$ of the sequence, via the logical rules of L, or (T-Intro), or (T-Elim), or (ω-Rule).

The set $\{\varphi \in \mathscr{L}_T^{\rightarrow} \mid$ there is a ω_{L}-derivation of $\varphi\}$ will be called ω_{L}-*logic*. ▲

Theorem 3.5.4.

No ω_{L}-logic yields Field's truth-set.

Proof sketch.

Let L be such that the resulting ω_{L}-logic is consistent, call such theory F. There is a positive inductive definition whose least fixed point is identical to F.[42] So, the computational complexity of F is less than or equal to Π^1_1, and thus cannot be Field's truth-set, by the previously mentioned result of [Welch 2009]. □

Some consider it to be a nice feature of a semantic construction that it can be obtained syntactically by a suitable infinitary theory:[43] the least Kripke fixed point is such a construction, since it can be defined by a simple ω-logic. This is not the case for Field's revision theory. Theorems 3.5.2 and 3.5.4 show how the high computational complexity of Field's revision theory that many were critical of has a significant impact on the possibility to characterize, and thus understand, the conditional connective it features.

3.5.4 Avenues of reply and primitivism about logical principles

In the foregoing arguments of this section, I have shown that some common ways of understanding what a conditional is are blocked for Field's conditional. I do not claim to have given an exhaustive discussion of all the possibly relevant readings of the conditional – this would be far beyond the scope of the present work. Nevertheless, the kinds of reading I have been discussing (model-theoretic/algebraic

[42] Folklore; see, for example, [McGee 1991].
[43] See for example [D.A. Martin 2011].

order-theoretic reading, characterization via introduction and elimination rules, and axiomatizations) are certainly the most relevant for formal theories of truth. So, it seems fair to conclude that the fact that Field's conditional does not fit into any of these paradigms is a serious problem for Field's approach, especially because, as I have argued, some understanding of the conditional should be specified in order to be in a position to accept or reject Field's theory. I do not exclude that there may be other ways to understand Field's conditional, that employ some other notion of meaning for logical constants, but the question of how to read Field's conditional is a pressing issue that currently has no satisfactory solution.

To address these difficulties, the Fieldian theorist could adopt a reply along the following lines: "it is OK to specify no interpretation for the conditional since, independently of that, several principles considered to be worth having can be consistently validated". I am not sure of how this reply can work. If the reply means something like "adopt any reading you like, still certain principles are recovered, and that's all that matters", we have a problem with the "adopt any reading you like" part, since many common and interesting readings of the conditional are barred by Field's theory, or not supported at best. Then we cannot adopt them, and the Fieldian theorist is challenged to come up with a new reading. If, instead, the reply means something like "it is OK to adopt no interpretation for the conditional *per se*, still certain principles are recovered, and that's all that matters", things are different. Here the Fieldian theorist expresses a sort of *primitivism* about logical principles: they embody irreducible aspects of the conditional and should be accepted, period. In the remainder of this subsection, I will refer to "principles" meaning "classically valid conditional principles", unless otherwise specified.

There are two main ways to interpret the primitivism I just described. One possibility is to argue that there is one specific set of principles that must be recovered. Let's call this position *specific primitivism*. In absence of a philosophically significant interpretation for the conditional, specific primitivism has unpleasant consequences. Nearly every theorist agrees that a theory of naïve truth should provide a classical interpretation of the truth-free part of the language. It is from *classical* semantics that naïve truth forces a deviation – hence the effort to recover as many classical principles as possible, or to approximate them at best. Suppose that a specific primitivist advocates the principles recovered by Field's theory, call them $\mathcal{Princ}$: what if another construction is found that does not validate all the principles in $\mathcal{Princ}$ but validates those in another set $\mathcal{Princ}^\star$, which has many principles in common with $\mathcal{Princ}$, and whose remaining ones look equally natural and fundamental? As Field's theory does not support the three major options for the meaning of the conditional considered in Subsections 3.5.1-3.5.3, and it does not provide a new interpretation for it, our theorist accepts some principles while being unable to use the meaning of the conditional to justify their acceptance. On what grounds, then, is our specific primitivist going to choose between $\mathcal{Princ}$ and $\mathcal{Princ}^\star$, or to argue for $\mathcal{Princ}$ over $\mathcal{Princ}^\star$? Choosing $\mathcal{Princ}$ over $\mathcal{Princ}^\star$ for no reason is clearly unacceptable.[44] Moreover, since we work in $\mathscr{L}_T^{\rightarrow}$ and we aim at recovering classical principles, many kinds of considerations to choose some principles over others are ruled out, e.g. those involving appeals to natural languages.

Alternatively, one could abandon specific primitivism for a weaker form, quantifying existentially over the principles to be recovered: "it is OK to adopt no interpretation for the conditional *per se*, still there exists a collection of principles that are recovered, and that's all that matters". Let's call this view *generic primitivism*. I don't have an argument against this position in general, also because it seems to be a fundamental view about what should we aim for in devising our theories of truth and it is difficult, and perhaps pointless, to argue in favor or against such fundamental views. Two such views can be dialectically articulated here, diverging on

[44] Especially in the area of theories of naïve truth, as recalled at the beginning of Section 3.5.

whether we want the addition of naïve truth to allow us to recover some principles or to preserve a certain understanding of the elements of our language. Donald Martin formulates a position not far from the latter view in the following terms:

> There is a question of *what theory* the construction yields a conservativeness proof for.[45] [...] I don't see how [Field's conditional] is *a* generalization; that is, I don't see *what* generalization it is supposed to be. In the end, all we are given is the model-theoretic construction and a (necessarily very partial) list of the laws and nonlaws. Contrast this with the connectives in the Kripke case. I would go so far as saying that Kripke's disjunction and conjunction *are* the classical disjunction and conjunction. ([D.A. Martin 2011], p. 345)

Generic primitivism, however, seems very difficult to defend when, as in Field's theory, we cannot combine it with a philosophically significant reading of the conditional, as this argument shows.

- Generic primitivism, by itself, only indicates to validate *at least one* set of principles, no matter *which*. This position is tremendously weak: a theory validating only the uninteresting law $(\varphi \wedge \psi) \to [(\varphi \to \chi_0) \to ((\chi_0 \to \chi_1) \to \chi_1)]$ satisfies the *desideratum* of generic primitivism. Moreover, a theory validating only the law $(\varphi \wedge \psi) \to [(\varphi \to \chi_0) \to ((\chi_0 \to \chi_1) \to \chi_1)]$ meets the requirements of generic primitivism just as a theory validating $\varphi \to \varphi$ and other apparently more interesting schemata. So, not only generic primitivism alone gives us no reasons to decide between two theories validating different sets of principles (as in the case $\mathcal{P}rinc$ vs. $\mathcal{P}rinc^{\star}$ from above), but it actually supports the conclusion that $\varphi \to \varphi$ is not more important than $(\varphi \wedge \psi) \to [(\varphi \to \chi_0) \to ((\chi_0 \to \chi_1) \to \chi_1)]$ and that in general no principle is more fundamental than another.

[45]Of course, this remark has to be understood for Field's sense of "conservativeness", namely model-theoretic conservativeness over ω-models of the ground language. See [Field 2008], p. 66.

- These undesirable consequences follow from generic primitivism *alone*: a philosophically significant reading for the conditional would give us a criterion to rule them out, but the Fieldian theorist cannot do this move.

- Since the consequences of generic primitivism just indicated are unacceptable, one must find a way to single out some principles as more important than others. Well, which are those principles and why one should choose exactly them? It is easy to see that we are drifting again toward the search for a kernel of "really fundamental" principles, and we will encounter again problems of the kind $\mathcal{Princ}$ vs. $\mathcal{Princ}^{\star}$ mentioned above: in the absence of a philosophically significant reading for the conditional, a generic primitivist has either to accept the bad consequences mentioned above, or to opt for an equally untenable specific primitivism.

A primitivist about principles seems to have no alternatives besides specific and generic primitivism: either she aims at recovering a specific set of principles, or an unspecified one. As both positions have intolerable consequences when no philosophically significant reading of the conditional is available, neither is an acceptable position for the Fieldian theorist.

Notice how focusing on the preservation of the meaning of connectives and quantifiers is a more neutral common ground: people disagreeing on which are the correct/best logical principles may agree on a more minimal notion of meaning for connectives and quantifiers. For example, the presence of different algebraic interpretations of the conditional does not generate a problem of the kind $\mathcal{Princ}$ vs. $\mathcal{Princ}^{\star}$, i.e. of which meaning is more fundamental: we seem to have an intuition according to which $\varphi \to \varphi$ is more fundamental than $(\varphi \wedge \psi) \to [(\varphi \to \chi_0) \to ((\chi_0 \to \chi_1) \to \chi_1)]$, but it seems unproblematic to say that no algebraic interpretation of the conditional is more fundamental than another – they are just different.

3.6 Determinateness

Field's theory preserves Kripke's theory for non-conditional sentences, since the least Kripke evaluation $\mathscr{K}_{\emptyset,\emptyset}$ is a subset of $\mathscr{U}$ and $\mathscr{F}$. So, for every $\varphi \in \mathscr{L}_T^{\rightarrow}$:

- if $\mathscr{K}_{\emptyset,\emptyset}(\varphi) = 1$, then $\mathscr{U}(\varphi) = 1^u$ and $\mathscr{F}(\varphi) = \mathbf{1}$,
- if $\mathscr{K}_{\emptyset,\emptyset}(\varphi) = 0$, then $\mathscr{U}(\varphi) = 0^u$ and $\mathscr{F}(\varphi) = \mathbf{0}$.[46]

In addition, the algebraic evaluation clauses that obtain in Kripke's construction for $\neg$, $\wedge$, $\vee$, $\forall$, and T are preserved, in a rather straightforward sense. Under the evaluation $\mathscr{F}$, indeed, the value of a negated sentence is given by a pointwise difference between $\mathbf{1}$ and the value of the negand, the value of a conjunction is the pointwise *minimum* of the values of the conjuncts, and so on. More formally:

Lemma 3.6.1.
For every $\varphi, \psi \in \mathscr{L}_T^{\rightarrow}$, and every $\chi(x) \in FOR_{\mathscr{L}_T^{\rightarrow}}$:

$$\mathscr{F}(\neg\varphi) = \langle 1 - r(\delta_0 + \alpha, \varphi) : \alpha < \gamma^+ \rangle.$$

$$\mathscr{F}(\varphi \wedge \psi) = \langle \min(r(\delta_0 + \alpha, \varphi), r(\delta_0 + \alpha, \psi)) : \alpha < \gamma^+ \rangle.$$

$$\mathscr{F}(\varphi \vee \psi) = \langle \max(r(\delta_0 + \alpha, \varphi), r(\delta_0 + \alpha, \psi)) : \alpha < \gamma^+ \rangle.$$

$$\mathscr{F}(\forall x \chi(x)) = \langle \inf\{r(\delta_0 + \alpha, \chi(t_n)) : \textit{for all } t_n \in TER_{\mathscr{L}_T^{\rightarrow}}\} : \alpha < \gamma^+ \rangle.$$

$$\mathscr{F}(T\ulcorner\varphi\urcorner) = \langle r(\delta_0 + \alpha, \varphi) : \alpha < \gamma^+ \rangle = \mathscr{F}(\varphi).$$

The similarities in the treatment of non-conditional sentences in Kripke's and Field's theories are evident comparing this result to Lemmata 2.4.3 and 2.5.5.

So, Field's theory inherits many features from Kripke's theory, but it is able to overcome some of the expressive difficulties emphasized in Section 2.5 thanks to its strong conditional. Nonetheless, some of the critical aspects of Kripke's theory

[46]Obviously, this remark works for the least Kripke evaluation $\mathscr{K}_{\emptyset,\emptyset}$ but it would not work for the least *C-gappy* Kripke evaluation $\mathscr{K}^+_{\emptyset,\emptyset}$, since the latter sends every conditional sentence to $1/2$.

remain unaddressed even in Field's framework. The remainder of this section is devoted to analyze some important improvements that Field's theory affords us over Kripke's theory, their costs and benefits, but also some noteworthy limitations that affect Kripke's approach and that are not tackled by Field's construction.

To introduce the topic of this section, consider again the liar sentence – clearly, however, the following considerations hold also for all the relevantly similar cases. As I have argued,[47] there is no way for Kripke's theory to address the lack of classical value of the liar sentence, nor to claim that it is neither true nor not true: the formalization available in Kripke's theory is $\neg T\ulcorner\lambda\urcorner \wedge \neg\neg T\ulcorner\lambda\urcorner$, which is just $\lambda \wedge \lambda$.[48] Still, there is clearly a sense in which λ *fails* to be true. If we had a stronger negation (such as the one I just used in English to say that λ *fails* to be true), we could express this fact in the object language, formalizing it with a sentence that is given the designated value by the semantics. Equivalently, we could keep the strong Kleene negation and try to express a notion that reinforces it, e.g. formalizing the idea that the liar sentence is *not determinately true*.

Now, Field's conditional can be used to express a notion of determinateness, formalizing and validating the claim that λ is not determinately true. In fact, the notion of determinateness has a crucial importance in Field's theory: not only it is the main tool to formalize and validate interesting semantic claims, especially about paradoxes, but it is also a paradigmatic case of Field's handling of *revenge paradoxes*.[49] As Field makes it clear, with its application to determinateness, "the new conditional can be used to show that the theory is not subject to «revenge problems»".[50] In the following, thus, I will focus on Field's approach to determinateness: this will contribute to the assessment of Field's theory and to identify

[47] See Subsection 2.5.1.
[48] Or $\lambda \wedge \neg\lambda$, or a conjunction of any two of λ and $\neg\lambda$ anyway.
[49] I will introduce revenge paradoxes in Subsection 3.6.2, when discussing them in relation to Field's theory.
[50] [Field 2003a], p. 140. Determinateness is investigated at length in [Field 2003a], [Field 2007] (Part Three), and [Field 2008] (Chapters 22 and 23).

some aspects of it that need to be addressed and improved.

3.6.1 Field's determinateness hierarchies: a sketch

In what follows, I will only consider Field's fine-grained evaluation $\mathscr{F}$ and its value space $\mathbf{F}$. Field's formalizes the notion of determinateness as an abbreviation that makes an essential use of the conditional.

Definition 3.6.2. (Field's determinateness operator)
For every $\varphi \in \mathscr{L}_T^{\rightarrow}$ let the operator $D(\varphi)$, for "determinately φ", be defined as the following abbreviation: $D(\varphi) := \varphi \wedge \neg(\varphi \rightarrow \neg\varphi)$. ▲

Thanks to Theorem 3.4.1, we can always introduce occurrences of T in every sub-sentence of a given sentence without altering its value. For this reason, it is not necessary to introduce a determinateness *predicate*, which can always be mimicked using the truth predicate, e.g. as $T\ulcorner D(\varphi)\urcorner$. So I will use expressions such as "determinately φ" and "φ is determinately true" (and the like) indifferently. The operator D possesses some features that make it apt to express the "semantic deficiency" of λ and other sentences.

Theorem 3.6.3. (Field)
For every $\varphi, \psi \in \mathscr{L}_T^{\rightarrow}$, the following holds:

If $\mathscr{F}(\varphi) = \mathbf{1}$, then $\mathscr{F}(D(\varphi)) = \mathbf{1}$.

If $\mathscr{F}(\varphi) \preceq \mathscr{F}(\neg\varphi)$, then $\mathscr{F}(D(\varphi)) = \mathbf{0}$. So, if $\mathscr{F}(\varphi) = \mathbf{0}$, $\mathscr{F}(D(\varphi)) = \mathbf{0}$.

If $\mathbf{0} \prec \mathscr{F}(\varphi) \prec \mathbf{1}$, then $\mathscr{F}(D(\varphi)) \prec \mathscr{F}(\varphi)$.

If $\mathscr{F}(\varphi) \preceq \mathscr{F}(\psi)$, then $\mathscr{F}(D(\varphi)) \preceq \mathscr{F}(D(\psi))$.

The behavior of D described by this theorem is quite intuitive. If a sentence has value $\mathbf{1}$, it is also determinately true. If a sentence has a value that is less than or equal to the value of its negation (in Field's ordering $\preceq$), it is not determinately true (and so $\mathscr{F}(\neg D(\varphi)) = \mathbf{1}$). If a sentence has a non-classical value (i.e. neither $\mathbf{0}$ nor $\mathbf{1}$), the sentence that declares it determinately true has a lower value. Finally, D respects the $\preceq$-ordering of values.

The determinateness operator gives the desired results on its intended targets, e.g. $\mathscr{F}(\neg D(\lambda)) = \mathbf{1}$, justifying the claim that Field's theory overcomes some expressive weaknesses of Kripke's theory. However, we can now formulate paradoxes that involve Field's determinateness. More specifically, we can define a sentence $\neg D(Tt_{\lambda_1})$, where $t_{\lambda_1} = \ulcorner \neg D(Tt_{\lambda_1}) \urcorner$, so that λ_1 and $\neg D(T\ulcorner \lambda_1 \urcorner)$ become provably equivalent (in weak arithmetical theories). Moreover, clearly $\mathscr{F}(\lambda_1) = \mathscr{F}(\neg D(\lambda_1)) \prec \mathbf{1}$, and the same holds reiterating $\neg D$ any number of times in front of the above sentences.

λ_1 brings us back to where we were with λ: there is a clear sense in which λ_1 *fails to be determinately true*. To capture this idea, Field suggests to *iterate* the determinateness operator: even if we cannot say that λ_1 is not determinately true, we can say that it is *not determinately determinately true*. And indeed $\mathscr{F}(\neg D(D(T\ulcorner \lambda_1 \urcorner))) = \mathbf{1}$. This triggers a further paradoxical sentence λ_2 that is equivalent to $\neg D(D(T\ulcorner \lambda_2 \urcorner))$, which is successfully treated by three iterations of the operator D, and so forth.

This process extends naturally into the transfinite, via the truth predicate: we could say that φ is determinately true at level ω if, for all natural numbers n, the sentence obtained prefixing φ with n occurrences of D is true. Formally, this is done defining a function $d(n, \ulcorner \varphi \urcorner)$ that yields the sentence resulting by prefixing φ with n iterations of D, and representing this function in the object language, similarly to the definition of McGee's sentence.[51] Field uses such mechanism to formalize iterations of the determinateness operator, indexed by ordinals, within

[51] See Section 1.3.

the object language: these iterations are highly infinitary, as they extend beyond the first non-recursive ordinal. In order to give a treatment of determinateness along these lines, two tasks should to be accomplished:

(I) to define a precise notion of iteration of D for some fragments of ordinals;

(II) to characterize the behavior of the iterations of D for the ordinals where they are defined.

Task (I) is performed by building *hierarchies of determinateness operators*. This construction, of remarkable technical complexity, is described in detail in [Field 2007]: I will give only a quick description, which will however suffice to analyze its more relevant aspects.[52]

Field defines large systems of ordinal notations to represent the iterations of D in the language. He constructs a set of functions $\mathcal{P}$, whose elements are called "paths", which associate with a fragment of ordinals some $\mathcal{L}_T^{\rightarrow}$-formulae to serve as their notations.[53] He then defines "path-dependent" hierarchies of iterations of D, where D is iterated along ordinals expressed in $\mathcal{L}_T^{\rightarrow}$ by some path. Given a path $p_0 \in \mathcal{P}$, the hierarchy of iterations of D dependent on p_0 works as follows:

- the 0-th iteration of D in front of φ yields φ;
- the $\alpha+1$-th iteration of D in front of φ expressed in p_0 yields the result of applying D to the result of the α-th iteration of D in front of φ in the notation for α provided by p_0;
- the β-th iteration of D in front of φ expressed in p_0, for β a limit ordinal, is equivalent to the infinite conjunction of all the results of the iterations of D

[52] The reader interested in the technical presentation of Field's determinateness hierarchies can see [Field 2007], sections 10-18 (pp. 109-134) and the Appendix (pp. 141-143). See also [Welch 2014].

[53] As Field notices ([Field 2008], p. 326, footnote 1), the "notations" he employs are not notation systems in the standard sense of Kleene (see [Kleene 1938] and [Rogers 1987] (Chapter 11)). This choice is due to the possibility that the language chosen as base language may not have enough closed terms to construct notations in Kleene's sense.

in front of φ indexed by ordinals smaller than β expressed in p_0 (they also have a notation in p_0, if β does).[54]

The dependence on notation makes the resulting hierarchies not very interesting, so Field defines one "path-independent" hierarchy of iterations of D as well. Path-independency is achieved by treating the ordinals which index the iterations of D in each path-dependent hierarchy *as variables*, and quantifying existentially over paths. So, for example, the α-th path-independent iteration of D in front of φ is the formalization in $\mathscr{L}_T^{\rightarrow}$ of the following sentence:

> "for some path p defined for a fragment of ordinals at least greater than α, the result of the α-th path-dependent iteration of D in front of φ depending on p is true".

For simplicity, I will use "Field's hierarchy" (or a similar expression) for the path-independent hierarchy, writing "$D^{\underline{\alpha}}$" for the α-th iteration of D in it.

As for the task (II) outlined above, let me formalize the notion of a *nicely-behaved iteration of D*.

Definition 3.6.4. [55]
Let's say that $D^{\underline{\alpha}}$ is *nice* if $\mathscr{F}(T\ulcorner D^{\underline{\alpha}}(0=0)\urcorner) = \mathbf{1}$. ▲

Non-nice applications of D do not even recognize arithmetical truths as determinate. By its construction, Field's hierarchy has non-nice iterations of D in it. In addition, it is a non-bivalent matter whether a fragment of the hierarchy is nice: there are iterations $D^{\underline{\beta}}$ in Field's hierarchy such that

(3.7) $$\mathscr{F}(T\ulcorner D^{\underline{\beta}}(0=0)\urcorner) \neq \mathbf{1} \text{ and } \mathscr{F}(\neg T\ulcorner D^{\underline{\beta}}(0=0)\urcorner) \neq \mathbf{1}.$$

[54] For completeness, I should add that Field considers several ways to define paths generating notations for ordinals, and there are crucial differences hinging on whether one maps ordinals to formulae of the full language $\mathscr{L}_T^{\rightarrow}$, to some proper subset of the formulae of the full $\mathscr{L}_T^{\rightarrow}$, or to formulae of the truth-free language $\mathscr{L}$. I will not go into details here, however, also because the more interesting treatment of iterations of D uses paths defined via all the formulae of the full $\mathscr{L}_T^{\rightarrow}$.

[55] This definition is a slight modification of the definition in [Field 2008] p. 330.

From fact (3.7), moreover, we have immediately the following:

$$(3.8) \qquad \not\models_F T\ulcorner D^{\underline{\beta}}(0=0)\urcorner \vee \neg T\ulcorner D^{\underline{\beta}}(0=0)\urcorner.$$

I collect in one theorem the main negative facts about Field's hierarchy.

Theorem 3.6.5. (Field)

1. *No nice $D^{\underline{\beta}}$ in Field's hierarchy is* bivalent*, i.e. such that for every $\varphi \in \mathscr{L}_T^{\rightarrow}$:*

$$\mathscr{F}(D^{\underline{\beta}}(\varphi)) = \mathbf{1} \textit{ iff } \mathscr{F}(\varphi) = \mathbf{1},$$
$$\mathbf{0} \textit{ iff } \mathscr{F}(\varphi) \prec \mathbf{1}.$$

2. *No nice $D^{\underline{\alpha}}$ in Field's hierarchy is* a limit*, i.e. such that for every $\varphi \in \mathscr{L}_T^{\rightarrow}$:*

$$\mathscr{F}(D^{\underline{\alpha}}(\varphi)) = \mathscr{F}(D^{\underline{\alpha+1}}(\varphi)).$$

It follows that no nice $D^{\underline{\alpha}}$ in Field's hierarchy is idempotent*, i.e. such that for every $\varphi \in \mathscr{L}_T^{\rightarrow}$, $\mathscr{F}(D^{\underline{\alpha}}(\varphi)) = \mathscr{F}(D^{\underline{\alpha}}(D^{\underline{\alpha}}(\varphi)))$.*

Proof.
Item 1 is proven via a liar-like reasoning. Suppose that $D^{\underline{\beta}}$ is nice and bivalent. Let λ_β be $\neg D^{\underline{\beta}}(Tt_\beta)$ for a term t_β such that $t_\beta = \ulcorner \neg D^{\underline{\beta}}(Tt_\beta)\urcorner$.

- If $\mathscr{F}(\lambda_\beta) = \mathbf{1}$, then $\mathscr{F}(D^{\underline{\beta}}(\lambda_\beta)) = \mathbf{1}$ and also $\mathscr{F}(D^{\underline{\beta}}(T\ulcorner\lambda_\beta\urcorner)) = \mathbf{1}$, but then $\mathscr{F}(\neg D^{\underline{\beta}}(T\ulcorner\lambda_\beta\urcorner)) = \mathbf{0}$, i.e. $\mathscr{F}(\lambda_\beta) = \mathbf{0}$, against our supposition.
- If $\mathscr{F}(\lambda_\beta) \prec \mathbf{1}$, then $\mathscr{F}(D^{\underline{\beta}}(\lambda_\beta)) = \mathbf{0}$ and also $\mathscr{F}(D^{\underline{\beta}}(T\ulcorner\lambda_\beta\urcorner)) = \mathbf{0}$, but then $\mathscr{F}(\neg D^{\underline{\beta}}(T\ulcorner\lambda_\beta\urcorner)) = \mathbf{1}$, i.e. $\mathscr{F}(\lambda_\beta) = \mathbf{1}$, against our supposition.

By item 1, for every nice $D^{\underline{\beta}}$ in Field's hierarchy there is at least one sentence $\varphi \in \mathscr{L}_T^{\rightarrow}$ such that $\mathscr{F}(D^{\underline{\alpha}}(\varphi)) \neq \mathbf{1}$ and $\mathscr{F}(D^{\underline{\alpha}}(\varphi)) \neq \mathbf{0}$. Then, by Theorem 3.6.3

and the definition of Field's hierarchy, $\mathscr{F}(D^{\underline{\alpha+1}}(\varphi)) \neq \mathscr{F}(D^{\underline{\alpha}}(\varphi))$. The claim about idempotent determinateness operators is immediate. $\square$

The above result shows that nice iterations of D in Field's hierarchy extend indefinitely and never converge to a single notion of determinateness that behaves as their limit or fixed point. Also, they never define a bivalent notion, according to which everything is determinately true or not.

The main positive results about Field's hierarchy are the following.

Theorem 3.6.6. (Field's hierarchy theorem)[56]

Let $D^{\underline{\beta}}$ belong to Field's hierarchy and be nice. Then, for every $\alpha < \beta$:

1. *For every $\varphi \in \mathscr{L}_T^{\rightarrow}$, $\mathscr{F}(D^{\underline{\beta}}(\varphi)) \preceq \mathscr{F}(D^{\underline{\alpha}}(\varphi))$.*

2. *For some $\psi \in \mathscr{L}_T^{\rightarrow}$, $\mathscr{F}(D^{\underline{\beta}}(\psi)) \prec \mathscr{F}(D^{\underline{\alpha}}(\psi))$.*

[Welch 2014] gives a sophisticated and compelling reconstruction of Field's hierarchy, providing a method to construct paths within Field's model. Welch shows that δ_0 (the least acceptable ordinal) is the upper bound of every α such that $D^{\underline{\alpha}}$ is in Field's hierarchy. In fact, if this were not the case, it would be possible to "diagonaliz[e] past the determinateness hierarch[y]" and formulate an "ineffable liar", namely a sentence "whose defectiveness is not encompassed by any [iteration of D] genuinely in [Field's hierarchy]".[57]

Theorem 3.6.7. (Welch's Ineffable Liars)[58]

Using iterations of D up through δ_0, *it is possible to define $\mathscr{L}_T^{\rightarrow}$-sentences λ_υ such that, for every $D^{\underline{\alpha}}$ in Field's hierarchy, $\mathscr{F}(D^{\underline{\alpha}}(\lambda_\upsilon)) \in \mathbf{E}$.*

[56]For a proof, see [Field 2008], p. 238.
[57][Welch 2014], p. 8.
[58]My formulation of Welch's result is a bit vague since making it precise would require to go through Welch's construction, but this is not necessary to understand the result.

Welch's limitative result may be thought to be the effect of a too encompassing quantification. After all, even if the hierarchy includes ordinals "very close" to δ_0, the iterations of D cease to be nice well before that stage – even if, by facts (3.7) and (3.8), we cannot isolate the fragment of the nice iterations in the hierarchy using Field's evaluation $\mathscr{F}$. Still, one may think that it could be possible to quantify over all and only the nice iterations of D and come up with an operator for *hyper-determinateness* equivalent to their infinite conjunction, to be added to the language. The semantics of such operator would be expressed in the meta-theory, but it may terminate the ever-increasing sequence of stronger and stronger iterations of D of the hierarchy. However, Field argues convincingly (albeit without a formal proof) that one of the following outcomes is the case:[59]

- The hyper-determinateness operator is non-nice for every $\mathscr{L}_T^{\rightarrow}$-sentence.

- It is not the case that the hyper-determinateness operator cannot be strengthened, and thus that terminates Field's hierarchy – this point can be reiterated for quantifications over nice applications of hyper-determinateness and stronger operators defined similarly.

3.6.2 Determinateness, revenge, and expressive limits

At the beginning of this section, I recalled that the treatment of determinateness is the keystone of Field's claim that his theory is "revenge immune".[60] Now, there is no universal consensus as to what counts as revenge paradox, but here I do not take a position on revenge paradoxes in general, I simply analyze some difficulties for Field's account of determinateness that are often seen as revenge problems.

[59] See [Field 2008], pp. 330-338.
[60] See page 90.

Revenge worries about determinateness

A formidable problem for Field's hierarchy is that it never yields a bivalent operator. The presence of such an operator in Field's framework would immediately generate an inconsistency, so it cannot even be added to the hierarchy.[61] Thus, Field's theory fails to account for an intuitive notion, according to which a given sentence is either absolutely determined or it fails to be so – one can replace "absolute" with her favorite adjective or periphrasis.[62]

We may follow [Leitgeb 2007] and formalize bivalent determinateness as a value-function defined in the meta-theory of Field's theory:

$$d_{bv}(\varphi) := \begin{cases} \mathbf{1} \text{ if and only if } \mathscr{F}(\varphi) = \mathbf{1} \\ \mathbf{0} \text{ if and only if } \mathscr{F}(\varphi) \neq \mathbf{1} \end{cases}$$

As Leitgeb emphasizes, the existence of exactly one function d_{bv} is a simple fact of many set theories. This includes (essentially all) the set theories that can be used as the *meta-theory* of Field's theory. Let's suppose, without loss of generality, that a theory including ZFC is used as meta-theory of Field's theory. By Theorem 3.6.5, d_{bv} cannot be *expressed* in $\mathscr{L}_T^{\rightarrow}$ consistently with Field's theory; in other words, there is no $\mathscr{L}_T^{\rightarrow}$-definable operator D^* (and *a fortiori* no predicate) such that:

$$\mathscr{F}(D^*(\varphi)) := \begin{cases} \mathbf{1} \text{ if and only if } \mathscr{F}(\varphi) = \mathbf{1} \\ \mathbf{0} \text{ if and only if } \mathscr{F}(\varphi) \neq \mathbf{1} \end{cases}$$

Clearly, an operator that behaves as D^* cannot be added consistently to $\mathscr{L}_T^{\rightarrow}$.

Alternatively, we can define a meta-theoretic bivalent notion of determinate-

[61] This follows immediately from the proof of Theorem 3.6.5.

[62] Equivalently, this problem could be recast as the lack of a negation that is stronger than the strong Kleene negation used in Field's theory. Some paraconsistent logicians argue for a negation that obeys some form of excluded middle; see for example [Beall 2009].

ness using Field's hierarchy. Consider the following sets:

$$\begin{aligned}
\mathcal{B}v^{+} &:= \{\varphi \in \mathcal{L}_T^{\rightarrow} \mid \mathscr{F}(D^{\underline{\alpha}}(\varphi)) = \mathbf{1}, \text{ for some nice } D^{\underline{\alpha}} \text{ in Field's hierarchy}\} \\
\mathcal{B}v^{-} &:= SENT_{\mathcal{L}_T^{\rightarrow}} \setminus \mathcal{B}v^{+} \\
&= \{\varphi \in \mathcal{L}_T^{\rightarrow} \mid \mathscr{F}(D^{\underline{\alpha}}(\varphi)) \neq \mathbf{1}, \text{ for every nice } D^{\underline{\alpha}} \text{ in Field's hierarchy}\}
\end{aligned}$$

Note that $\mathcal{B}v^{+}$ and $\mathcal{B}v^{-}$ are well-defined and consistent with facts (3.7) and (3.8).[63] Analogously to the case of the meta-theoretic value-function d_{bv}, by Theorem 3.6.5 no operator can express membership in $\mathcal{B}v^{+}$ or in $\mathcal{B}v^{-}$.[64]

The positive part of d_{bv} or of $\mathcal{B}v^{+}$ offers also a meta-theoretic, partial characterization of (the positive part of a) *limit* and *idempotent* notion of determinateness. A full, meta-theoretic characterization is given by the following value-function.

$$d_{lm}(\varphi) := \mathbf{v} \text{ if and only if } \mathscr{F}(D^{\underline{\alpha}}(\varphi)) = \mathbf{v} \text{ for every nice } D^{\underline{\alpha}} \text{ in Field's hierarchy.}$$

Again, d_{lm} is well-defined and consistent with facts (3.7) and (3.8), but it cannot be expressed in $\mathcal{L}_T^{\rightarrow}$ consistently with Field's theory.

The meta-theoretic characterizations of some aspects of the notion of determinateness offered by d_{bv}, by $\mathcal{B}v^{+}$ and $\mathcal{B}v^{-}$, or by d_{lm} are simple and clear. Some authors, e.g. Leitgeb, take these facts as evidence for the conclusion that "Field's

[63]That $\mathcal{B}v^{+}$ and $\mathcal{B}v^{-}$ are well-defined is immediate from Theorems 3.6.3 and 3.6.6. In fact, excluded middle fails in claims of the form "$D^{\underline{\alpha}}$ is nice or $\neg D^{\underline{\alpha}}$ is nice" in the sense that their formalization in $\mathcal{L}_T^{\rightarrow}$ does not always receives value **1** by Field's evaluation $\mathscr{F}$, and we cannot consistently define the set of all and only the nice iterations of $D^{\underline{\alpha}}$ *within Field's model*. However, we can use classical logic to make claims about Field's hierarchy *in its classical meta-theory*.

[64]Both the theories I develop in Chapter 4 and in Chapter 5 are able to express consistently all *positive* meta-theoretic assertions of determinateness (for the notions of determinateness available to those theories). More explicitly, both those theories are able to define consistently a *unique*, non-hierarchical operator D^{+} such that $D^{+}(\varphi)$ takes the designated value if and only if φ is declared determinate by the theory (for their notions of determinateness). What those theories cannot do is to express an operator D^{-} such that $D^{-}(\varphi)$ takes the designated value if and only if φ is not declared determinate by the theory – which, in those theories, is different from being declared *not determinate*. Field's theory, by contrast, is not even able to represent all positive meta-theoretic assertions of determinateness, namely $\mathcal{B}v^{+}$ alone, since it is impossible to quantify over all the nice iterations of D in Field's hierarchy. The same point applies to $\mathcal{B}v^{-}$ alone.

theory is affected by a revenge problem *with respect to the classical metatheory*".[65] This seems a fair conclusion to me: given the adoption of a classical meta-theory, this kind of revenge paradox is a price that non-classical theories have to pay, and it is reasonable to expect that the non-classical theorist should learn to live with it.[66] Of course, this does not limit in any way the interest of non-classical theories formulated in a classical meta-theory: as there are many non-classical solutions to truth-theoretic paradoxes (and some are better than others), there are also many ways in which the damage of revenge paradoxes can be minimized (and some ways are arguably better than others). The existence of limitative results, it seems to me, should not prevent one from developing theories of truth, determinateness, and other semantic notions that come closer and closer to recover the unexpressible ideal notions that the meta-theory gives.[67]

The attitude just sketched towards revenge, however, is not shared by Field.[68] He shows, in fact, to favor two kinds of replies to revenge problems. One way is to look for a non-classical theory of classes that is strong enough to produce a non-classical meta-theory in which, arguably, the problematic notions that give rise to revenge paradoxes are not definable. This line is suggested also in [Leitgeb 2007] but it presents considerable difficulties, several of which are explored in [Field, Lederman, and Øgaard (forthcoming)]. The second reply considered by Field, instead, is developed entirely within the context of classical meta-theories, and consists in denying that the problems above constitute real revenge paradoxes.

[65] [Leitgeb 2007], p. 105.

[66] The inexpressibility of bivalent determinateness, after all, is not a special defect of Field's approach, but it affects every theory that satisfies some rather minimal requirements. Suppose that one has a semantic theory of truth for $\mathscr{L}_T^{\rightarrow}$ that respects intersubstitutivity of truth. Let $\mathscr{E}$ be the resulting evaluation function, $\mathbf{O}$ be the set of designated values, with $\mathbf{d_0} \in \mathbf{O}$ and $\mathbf{a_0} \notin \mathbf{O}$. Let $\neg$ be interpreted so that $\mathscr{E}(\neg\varphi) \in \mathbf{O}$ iff $\mathscr{E}(\varphi) \notin \mathbf{O}$. A liar-like contradiction ensues from any operator expressing the following value-function:

$$d_{\mathscr{E}}(\varphi) := \begin{cases} \mathbf{d_0}, \text{ iff } \mathscr{E}(\varphi) \in \mathbf{O} \\ \mathbf{a_0}, \text{ iff } \mathscr{E}(\varphi) \notin \mathbf{O} \end{cases}$$

[67] The theories developed in Chapters 4 and 5 are in line with this attitude.

[68] And others, of course: see for example [Priest 1990].

The first kind of reply is really interesting, and is currently undergoing a lot of investigation;[69] however, its tenability and its consequences cannot be fully assessed now, as non-classical meta-theories of the kind that would be needed for it are not available yet. The second kind of reply, on the other hand, has been developed extensively by Field in [Field 2007], and the arguments given in its favor have also an independent conceptual and methodological interest: for these reasons, the second reply is analyzed now.

Trying to avoid revenge

Field's reply is designed to show that it is simply not a shortcoming of his approach that it does not yield a bivalent, a limit, or an idempotent notion of determinateness. For simplicity, the following considerations will focus on bivalent determinateness, but they extend to the limit or idempotent case.[70] Of course, Field is well aware of functions such as d_{bv} or sets such $\mathcal{B}v^{+}$ and $\mathcal{B}v^{-}$. Despite them, however, the notion of bivalent determinateness is not "ultimately intelligible when examined closely",[71] hence there is no revenge problem concerning it.

This conclusion is sometimes discarded quickly as wrong, with a reasoning of the following kind: the notions that Field claims to be unintelligible cannot be such, since they are used to articulate his view.[72] But this response misfires, as it could only work if bivalent determinateness were deemed *unconceivable* (or the like): while an unconceivable notion cannot be used to articulate a view, not even

[69] See, e.g., [Bacon 2013b].

[70] The choice to focus on bivalent determinateness here may seem unfaithful to Field's position, in view of verdicts such as the following one: "*there can be no truth-like predicate for which excluded middle can be assumed*" ([Field 2007], p. 89, emphasis in the original). Nevertheless, Field himself acknowledges that "the conviction that there *must* be truth-like predicates obeying excluded middle is one primary source of revenge worries" (*ibidem*): this, together with Leitgeb's work, shows that it would be inappropriate to exclude bivalent determinateness from general considerations about revenge problems involving determinateness.

[71] [Field 2007], p. 119.

[72] See for example [Priest 1990]: "[...] the dialetheist can argue *ad hominem* against consistent theorists that the notions they declare to be ineffable are intelligible, because they themselves use these very notions to explain their views."

as a negative standard, a notion that is not ultimately intelligible clearly can. Field's claim cannot be discarded so quickly. He offers a carefully devised argument for it, which needs a close scrutiny.[73]

Field's argument

(A) The starting point of Field's argument is the thesis that truth is a *model-independent notion*. The notion of truth, in fact, is characterized by the (T)-Schema, whose instances can be expressed in $\mathscr{L}_T^{\rightarrow}$ without appeal to model-relative notions, such as "having a/the designated value in a certain model/collection of models".[74] Contrast this, e.g., with the *semantic* rules (T-Intro) and (T-Elim): they give a model-dependent characterization of truth. Determinateness is also characterized by model-independent principles: even if Field is less explicit about them, determinateness "obey[s] the analogs of (T-Elim) and (T-Intro)"[75] and object-language versions of the principles encoded in Theorem 3.6.3.[76]

On the other hand, notions as "truth-in-the-model" and "determinate-truth-in-the-model" are mere surrogates of truth or determinate truth: often, in fact, true and determinately true claims turn out to be not true-in-the-model nor determinately-true-in-the-model, for many natural models.[77] Model the-

[73] See especially [Field 2007], §§7-8, pp. 102-109, §13 pp. 120-123, and §§19-20, pp. 134-141. For some other considerations on this argument, see [Scharp 2007], pp. 280-309.

[74] One could also use a version of intersubstitutivity of truth formulated with biconditionals.

[75] [Field 2007], p. 88.

[76] Namely:

(a) from φ, infer $D(\varphi)$;

(b) from $\varphi \rightarrow \neg\varphi$, infer $\neg D(\varphi)$;

(c) $D(\varphi) \rightarrow \varphi$;

(d) from $\varphi \rightarrow \psi$, infer $D(\varphi) \rightarrow D(\psi)$.

Item (c) should be strengthened slightly to match the third item in Theorem 3.6.3. Note that since some of these principles are in rule form, by the previous observation, they cannot be semantic inferences (otherwise model-independency would be lost).

[77] Field brings the following case in support of the latter claim. A "highly natural" model for the language of ZFC is "the homophonic model whose domain consists of all non-sets together with all

ory is instrumental to prove the consistency of model-independent principles, but the understanding of truth and determinateness is not mediated by model theory, nor conveyed by meta-theoretic definitions: we cannot "extrapolate an understanding of a model-independent notion like truth or determinate truth" from their model-dependent surrogates. It follows that meta-theoretic, model-dependent surrogates of bivalent determinateness (e.g. d_{bv} or $\mathcal{B}v^{+}$ and $\mathcal{B}v^{-}$) do not make understandable the notion they allege to represent. So, there is no revenge problem about meta-theoretic bivalent determinateness.

(B) To vindicate revenge-immunity, thus, it remains only to check that no revenge paradox arises *within* the theory. But Theorems 3.6.5 and 3.6.6 show that no operator for bivalent determinateness can be defined in the language, and we have seen at the end of Subsection 3.6.1 that hyper-determinateness does not yield a nice bivalent notion of determinateness either. So, as Field puts it, "there is *no way to generate an understanding* of a notion of [bivalent determinateness] from the hierarchy of determina[teness] notions that we have in the language".[78]

(C) By **(A)** and **(B)**, bivalent determinateness is unintelligible (as it is neither meta-theoretically nor intra-theoretically intelligible), so there is no revenge paradox involving it. This strategy can be applied to any notion that purportedly originates revenge worries, making Field's theory revenge-immune.

Evaluation of the argument

Field's argument, despite its interest and elegance, must ultimately be rejected.

---

sets of rank less than the first inaccessible cardinal" ([Field 2007], p. 104). Call this model $\mathscr{M}$. In order to build $\mathscr{M}$, we need the existence of at least one inaccessible cardinal. So, the sentence "there are inaccessible cardinals", which can be formalized in the language of ZFC, is true and determinately true, yet neither "true-in-$\mathscr{M}$" nor "determinately-true-in-$\mathscr{M}$".

[78] [Field 2007], p. 135, my emphasis.

I do not believe that there is something wrong with the argument in itself, namely in its development from its premisses to the conclusion. The argument, however, makes some tacit assumptions that have unacceptable consequences for our intelligibility standards: these assumptions must be discarded, but without them the argument does not seem to stand.

Field's argument declares that bivalent, limit, and idempotent determinateness (or combinations of the three) are not intelligible due to their model-dependent formulation. But object-linguistic, model-independent versions of bivalence, limitness, or idempotency for Fieldian determinateness are easy to formulate:

$$(\mathrm{Biv})^o \quad D^{\underline{\alpha_0}}(\varphi) \vee \neg D^{\underline{\alpha_0}}(\varphi),$$

$$(\mathrm{Lim})^o \quad D^{\underline{\beta_0}}(\varphi) \leftrightarrow D^{\underline{\beta_0+1}}(\varphi),$$

$$(\mathrm{Id})^o \quad D^{\underline{\gamma_0}}(\varphi) \leftrightarrow D^{\underline{\gamma_0}}(D^{\underline{\gamma_0}}(\varphi)),$$

where α_0, β_0, and γ_0 have a notation in the ordinal notation adopted.

Of course, $(\mathrm{Biv})^o$, $(\mathrm{Lim})^o$, and $(\mathrm{Id})^o$ are not formulated very clearly, since I have not specified the ordinals α_0, β_0, and γ_0. Nevertheless, they suffice to show that one can give model-independent characterizations of bivalent, limit, idempotent determinateness: it is enough to combine $(\mathrm{Biv})^o$, $(\mathrm{Lim})^o$, $(\mathrm{Id})^o$, or more than one of them with Field's other principles governing determinateness.

With Theorem 3.6.5 we have seen that bivalent, limit, or idempotent determinateness do not obtain in Field's theory, but this is not an impediment to *formulate* object-linguistic principles stating the corresponding conditions for determinateness, such as $(\mathrm{Biv})^o$, $(\mathrm{Lim})^o$, and $(\mathrm{Id})^o$. Such principles are model-independent characterizations of bivalent, limit, or idempotent determinateness just as the (T)-Scheme is a model-independent characterization of truth. But then, one could think that, by the lights of Field's own argument, they make the corresponding notions sufficiently intelligible, and we have a revenge problem involving them.

Field cannot accept this conclusion, so he has to find a way to ban $(\mathrm{Biv})^o$, $(\mathrm{Lim})^o$, and $(\mathrm{Id})^o$. It seems difficult, however, to ban $(\mathrm{Biv})^o$, $(\mathrm{Lim})^o$, and $(\mathrm{Id})^o$ on syntactic grounds: they are perfectly ordinary schemata of $\mathscr{L}_T^{\rightarrow}$, and they are well-formed, since α_0, β_0, and γ_0 have a notation in $\mathscr{L}_T^{\rightarrow}$. Since syntactic features are not useful here, we could resort to semantic considerations. We see immediately that there is no known model for $(\mathrm{Biv})^o$, $(\mathrm{Lim})^o$, or $(\mathrm{Id})^o$ and (at least several of) the principles that Field's theory validates.[79] But this exhausts the semantic considerations that we can make: combinations of $(\mathrm{Biv})^o$, $(\mathrm{Lim})^o$, or $(\mathrm{Id})^o$ and (at least several of) the principles targeted by Field have no known models so, even if such models existed, we couldn't talk about them now, as they are not known.

It is unlikely to find other kinds of considerations (besides syntactic and semantic ones) to be used to conclude that $(\mathrm{Biv})^o$, $(\mathrm{Lim})^o$, and $(\mathrm{Id})^o$ do not characterize model-independently bivalent, limit, or idempotent determinateness. As a consequence, the absence of known models of $(\mathrm{Biv})^o$, $(\mathrm{Lim})^o$, $(\mathrm{Id})^o$ and other principles targeted by Field is the only peculiarity of the former, and hence the only fact that we can use to claim that $(\mathrm{Biv})^o$, $(\mathrm{Lim})^o$, and $(\mathrm{Id})^o$ do not generate understandable notions. Field's argument, thus, needs to incorporate the following postulate.

(UM) For a semantic notion to be intelligible it is not sufficient to formulate model-independent principles governing it: such principles must have a model. Finding models for a model-independent notion is a necessary condition to understand it. In a slogan: "*Understanding* requires *Models*"

The Fieldian theorist should endorse **(UM)** since, as seen, denying it would make intelligible notions that Field needs to make unintelligible. Of course, she would not apply **(UM)** to $(\mathrm{Biv})^o$, $(\mathrm{Lim})^o$, and $(\mathrm{Id})^o$ alone, but to these principles together with intersubstitutivity of truth and at least some other principles that Field

[79] While I am writing, at least.

claims to be desirable. I will not be specific here, as this does not seem to obscure the main conceptual and methodological points of the discussion.

(UM) is not *prima facie* absurd. It filters out (Biv)^o, (Lim)^o, and (Id)^o nicely without declaring unintelligible notions that are obviously intelligible even if Field does not accept them: exclusion negation, for example, has dialethic models that validate also several of the principles that Field accepts, including intersubstitutivity of truth. In addition, **(UM)** may be defensible in some non-semantic cases: for example, since all models of second-order Peano Arithmetic are isomorphic, it may be argued (perhaps from a structuralist perspective) that one does not understand whatever arithmetical notions come from second-order Peano Arithmetic unless she has in mind a model having the characteristics common to every model.

However, applying **(UM)** to truth or determinateness would impose unacceptable revisions to our intelligibility standards. Take the case of truth, for example. Many believe that Tarski biconditionals capture the most natural and intuitive trait of truth. By **(UM)**, if one understands Tarski biconditionals, she has a model of them. So, **(UM)** entails that Tarski biconditionals were not intelligible when formulated for the first time, since, presumably, there was no model for them yet (ironically, this includes Tarski's formulation, whether it was the first formulation or not) – one should then wonder why some people bothered to formulate an unintelligible principle.[80] Presumably, moreover, no speaker of English who is not a logician familiar with non-classical theories and who accepts the very few classical principles involved in a derivation of a contradiction from Tarski biconditionals understands the latter, despite the appealing idea that they are natural and intuitive – unless we are willing to call "natural and intuitive" something that many people do not understand. Similar considerations show that **(UM)** entails untenable standards

[80]Of course, I refer to the inconsistency of Tarski biconditionals *with a sufficiently expressive background theory to serve as syntax and some logical principles*. I am making the historically plausible assumption that no model of Tarski biconditionals formulated in and for a suitably expressive language was available when Tarski biconditionals were formulated for the first time.

of intelligibility for the notion of determinateness endorsed by Field.

The Fieldian theorist may adopt a subtler distinction to save **(UM)**: there is a "shallow" sense of intelligibility in which we all understand semantic notions (characterized by model-independent principles), but there is also a "deep" sense in which understanding them requires a model, and only the deep sense is relevant for freedom from paradoxes. Let's consider again the notion of truth characterized by the (T)-Scheme. The shallow/deep distinction would make this notion of truth close to notions such as *black hole*: in a "shallow" sense many people understand what a black hole is, but in a "deep" sense only specialists do. The shallow/deep distinction, however, is misguided for naïve truth or determinateness. Indeed, it is hard to believe that only deeply intelligible notions are relevant to decide whether a certain paradox arises or not. Before the development of modern theories that validate the (T)-Scheme, for example, the notion of truth characterized by the (T)-Scheme was only shallowly intelligible: still, the (T)-Scheme was used in deriving contradictions from the liar sentence, and it is hardly maintainable that the liar paradox should not have been considered an authentic paradox. But then, if the (T)-Scheme generated authentic paradoxes also when it was only shallowly intelligible, why $(\text{Biv})^o$, $(\text{Lim})^o$, and $(\text{Id})^o$ do not generate authentic paradoxes?

Finally, notice that reversing **(UM)** and making the existence of a model a *sufficient* condition for intelligibility (whether shallow or deep) would not help Field's argument, which would still need a reason to rule out the possibility that some model-independent principles with no models are nonetheless intelligible.

In conclusion, the absence of known models for $(\text{Biv})^o$, $(\text{Lim})^o$, or $(\text{Id})^o$ does not support the thesis that bivalent, limit, or idempotent determinateness are unintelligible. Since the absence of known models is the only trait that tell $(\text{Biv})^o$, $(\text{Lim})^o$, and $(\text{Id})^o$ apart from the other semantic notions under discussion, that are characterized model-independently and accepted as intelligible, we have no basis to declare that bivalent, limit, or idempotent determinateness are not perfectly intelligible by the very standards employed in Field's argument. As a consequence, these notions give rise to authentic revenge paradoxes for Field's theory.

Approximations?

Since the attempt to shun revenge problems via Field's argument fails, we should turn to the more modest attitude discussed towards the beginning of Subsection 3.6.2, i.e. we should check how close Field's theory comes to bivalent determinateness and what are the relations between this unreachable ideal and Field's hierarchy. A natural option from this point of view is to say that Field's nice determinateness operators *approximate* bivalent determinateness. In fact, for every sentence φ s.t. $\mathscr{F}(\varphi) = \mathbf{1}$, we have $\mathscr{F}(D^{\underline{\alpha}}(\varphi)) = \mathbf{1}$ for every nice $D^{\underline{\alpha}}$ in the hierarchy. Moreover, for every ψ getting value different from $\mathbf{1}$, there is a nice $D^{\underline{\beta}}$ "high enough" in the hierarchy such that $\mathscr{F}(D^{\underline{\beta}}(\psi)) = \mathbf{0}$.

Speaking of approximations often involves the idea that the approximating notions behave as expected in many cases, leaving out a (hopefully small) residual. It seems accurate to say that, as the nice iterations of $D^{\underline{\alpha}}$ grow, they approximate bivalent determinateness better and better. It seems also fair to say that the residual left by an operator $D^{\underline{\alpha}}$ is constituted by those sentences for which $D^{\underline{\alpha}}$ does not behave bivalently. The residual of every $D^{\underline{\alpha}}$ can be defined in the meta-theory: call it $R_{D^{\underline{\alpha}}}$. It can be shown that if $D^{\underline{\alpha}}$ and $D^{\underline{\beta}}$ are nice and $\alpha < \beta$, then $R_{D^{\underline{\alpha}}} \supsetneq R_{D^{\underline{\beta}}}$. So, Field's hierarchy can be seen as a hierarchy of approximations to d_{bv}. However, even if every operator leaves a residual, there is no "overall" residual. So, it is problematic to say that Field's hierarchy itself constitutes an approximation of bivalent determinateness, unless one accepts the idea that the approximation leaves no residual, while failing nonetheless to deliver the non-approximated notion.

In fact, the most accurate paradigm for approximation embodied by Field's hierarchy is not that of a partial coverage of some cases, but that of a series that never converges to a limit.[81] The two paradigms, however, present marked differences. In the case of the partial coverage, we recover the approximated notion *modulo* a residual of non-covered cases; if the approximation is good, moreover, there will be a justification to explain the non-covered cases, and they will be as few as possible. What about the diverging series case? This paradigm entails that no single element of the hierarchy is good enough: whenever any such element fails to behave bivalently on a given sentence, another element further up in the hierarchy succeeds. So, we cannot be satisfied with any given element of the hierarchy. And, of course, the whole hierarchy itself cannot be replaced by a single operator expressible in the object language, as the previous results and discussion make clear.

These difficulties cast serious doubts on the fruitfulness of Field's hierarchy and of a hierarchical approach to determinateness in general. This prompts us to come up with a different treatment of determinateness, that avoids the weakness of Kripke's theory but also the unnatural, complicated, and ultimately inadequate hierarchical framework.[82]

3.7 Some more Kripkean limitations

There are more difficulties of Kripke's theory that are not overcome in Field's theory: e.g., the unsatisfactory treatment of the truth-teller given by Kripke's theory (see Subsection 2.5.1) is inherited by Field's theory. As in Kripke's theory, τ (and relevantly similar cases) can be given more than one value in Field's theory, and its value depends only on the Kripke fixed points adopted at revision stages. Any value that Field's theory can assign to truth-teller-like cases conflates their treatment with other sentences that *must* have that value, while the former merely *can*. Similar

[81] See [Leitgeb 2007] for a precise formulation of this paradigm about Field's hierarchy.
[82] Alternatives treatments of determinateness will be given in Sections 4.3 and 5.4.

remarks go for the sentences formalizing the claims that truth-teller-like cases are not determinately true: the value assigned to those sentences depends only on the chosen Kripke fixed points. No single evaluation built with Field's construction can account for the fact that truth-teller-like cases can take any value, which can only be explained via a collection of evaluations, just as in Kripke's theory.

3.8 Summary of the chapter

In this chapter, I gave a critical analysis of Field's theory, focusing on the meaning of Field's conditional, on the status of the principles recovered by his theory, and on determinateness and revenge paradoxes.

In Section 3.1, I introduced $\mathscr{L}_T^{\rightarrow}$ (Subsection 3.1.1) and discussed the relations between Field's approach and FIELD'S PROGRAM (Subsection 3.1.2). I presented Field's theory informally in Section 3.2 and more rigorously in Section 3.3, reviewing the main facts about it in Section 3.4. Section 3.5 motivated the question of how to understand Field's conditional: Subsection 3.5.1 analyzed the difficulties of seeing Field's conditional as a tool to compare semantic values, Subsection 3.5.2 dealt with the problematic conditional introduction of Field's theory, and Subsection 3.5.3 explored the impediments to axiomatizing Field's theory. Subsection 3.5.4 addressed some avenues of reply to the above difficulties and their costs. Section 3.6 discussed Field's treatment of determinateness, its functioning (Subsection 3.6.1), and revenge problems (Subsection 3.6.2). Section 3.7 recalled some Kripkean limitations inherited by Field's theory, concerning truth-teller-like cases.

The next chapter presents a new theory of truth that goes some way towards addressing FIELD'S PROGRAM, avoiding some of the difficulties of Field's approach.

Part II

Some New Semantic Theories of Naïve Truth

Chapter 4

Adding a Conditional to Kripke's Theory

4.1 Introduction

In Section 2.5, I analyzed the main expressive limitations of Kripke's theory. As argued, such limitations hinder significantly the efficacy of Kripke's account of the naïve behavior of truth, which may be improved by adding a good conditional to Kripke's theory. Field made fundamental contributions to this project, providing constructions that endow Kripke's theory with a strong conditional connective, validating several classically valid conditional principles together with intersubstitutivity of truth. Field's revision theory is probably the best theory he advanced. These achievements are of enormous technical importance; nevertheless, Field's theory falls short of providing a philosophically significant reading for the conditional, as I argued in Section 3.5. Such problem should not be underestimated: without a philosophically significant interpretation for the conditional, we are in no position to really understand and assess Field's contribution, to justify the acceptance of the principles that his theory recovers, and to accept his overall account of the semantics of $\mathscr{L}_T^{\rightarrow}$. These difficulties, in turn, transfer to the account of the notion of truth itself and to the paradoxes that affect it.

As mentioned in Section 2.6, one of the main purposes of this work is to address FIELD'S PROGRAM via a well-behaved conditional, that is interesting for the truth-theorist, simple to characterize and to interpret, and that plays a well-defined role in the theory of truth. The criteria I adopt for the good behavior of connectives and quantifiers, **(W1)**, **(W2)**, and **(W3)**, were elaborated and defended in Section 2.7.[1]

To motivate the strategy to address FIELD'S PROGRAM I will articulate in this chapter, let me return for a moment to the basics of Kripke's theory. Kripke's construction is based on strong Kleene semantics.[2] Now, the strong Kleene evaluation scheme has a weak conditional, but it makes its other connectives and quantifiers well-behaved. Interpreted as in strong Kleene semantics, negation, conjunction, disjunction, and the universal quantifier have a behavior that approximates closely the classical one, and a simple and conceptually relevant algebraic reading. As mentioned in Subsection 3.5.1, strong Kleene semantics offers an interesting reading of sentences whose main operator is $\neg$, $\wedge$, $\vee$, or $\forall$: they provide information on order-theoretic relations concerning the values of their immediate proper sub-sentences. The strong Kleene conditional, however, does not enter this picture, and (I argued) neither does Field's conditional. An appealing possibility to address

[1]I recall them here for the reader's convenience.

(W1) Every connective and quantifier has a clear and simple *meaning*. For example, it should be possible to describe the meaning of every connective and quantifier via some simple algebraic rule (given by the semantics), or to axiomatize adequately the $\mathscr{L}_T^{\rightarrow}$-sentences validated by the semantics, or to characterize every connective and quantifier via suitable introduction and elimination rules (validated by the semantics).

(W2) Every connective and quantifier as interpreted by the semantics should be closed under suitable introduction and elimination rules, that approximate the classical ones.

(W3) Conferring a meaning on the connectives and quantifiers, as specified in **(W1)**, should allow us to render (as much as possible) the intuitions we aimed at capturing with the development of our semantic theory.

[2]Recall that I do not consider the other evaluation schemes considered by Kripke (weak Kleene logic and supervaluations) since, as explained in Section 2.2 and at the beginning of Section 2.3, they are not very relevant for the problems addressed in the present work.

FIELD'S PROGRAM, thus, would be to find an evaluation scheme that supplements strong Kleene semantics with a conditional that can be interpreted as a tool to compare values, in the sense of Subsection 3.5.1. An obvious candidate for this role is *Łukasiewicz 3-valued semantics*, since its evaluation scheme includes the strong Kleene scheme and it features a primitive conditional which could fulfill the desired role.[3] A Łukasiewicz 3-valued conditional sentence, in fact, receives value 1 in all the cases in which a strong Kleene material conditional receives value 1 *plus* in the case where both its antecedent and its consequent have value $1/2$. I recall here every Łukasiewicz evaluation scheme, whether finitely or infinitely valued.[4]

Definition 4.1.1. (Łukasiewicz value spaces and evaluation schemes)
A *Łukasiewicz value space* V is either any set $\{0, 1/n+1, \ldots, n/n+1, 1\}$, where n is an odd positive integer,[5] or the real interval $[0,1]$. $|V|$ is the cardinality of V. The value functions of Łukasiewicz semantics are independent from $|V|$.
The *Łukasiewicz evaluation scheme* for $\mathscr{L}_T^{\rightarrow}$ is described by the functions $\mathscr{E}_{Ł} : SENT_{\mathscr{L}_T} \longmapsto V^*$ (where V^* is an arbitrary Łukasiewicz value space) that respect the following conditions: for every $\varphi, \psi \in \mathscr{L}_T$, and every $\chi(x) \in FOR_{\mathscr{L}_T}$,

$$\begin{aligned}
\mathscr{E}_{Ł}(\neg\varphi) &= 1 - \mathscr{E}_{Ł}(\varphi), \\
\mathscr{E}_{Ł}(\varphi \wedge \psi) &= \min(\mathscr{E}_{Ł}(\varphi), \mathscr{E}_{Ł}(\psi)), \\
\mathscr{E}_{Ł}(\varphi \vee \psi) &= \max(\mathscr{E}_{Ł}(\varphi), \mathscr{E}_{Ł}(\psi)), \\
\mathscr{E}_{Ł}(\varphi \rightarrow \psi) &= \min[1, (1 - \mathscr{E}_{Ł}(\varphi) + \mathscr{E}_{Ł}(\psi))], \\
\mathscr{E}_{Ł}(\forall x \chi(x)) &= \inf\{\mathscr{E}_{Ł}(\chi(t_n)) : t_n \in TER_{\mathscr{L}_T}\}.
\end{aligned}$$

▲

[3] See Definition 1.4.4.
[4] For Łukasiewicz logics and semantics, see [Malinowski 1993], [Gottwald 2001], [Urquhart 2002], [Hänle 2002].
[5] This allows us to evaluate the liar sentence and to assign a value to a sentence exactly when we are able to assign a value to its negation (see [Field 2008], p. 86).

It is well-known, however, that not only Łukasiewicz 3-valued semantics but in general every evaluation for $\mathscr{L}_T^{\rightarrow}$ based on a Łukasiewicz semantics does not mingle well with weak naïveté – let alone with stronger forms of naïveté.

Theorem 4.1.2. (Folklore (first half); Restall (second half))[6]

There is no total, n-valued function $\mathscr{E}_{Ł}^{n} : SENT_{\mathscr{L}_T^{\rightarrow}} \longmapsto V_{Ł}^{n}$ *s.t.:*

1. $\mathscr{E}_{Ł}^{n}$ *agrees with* $\mathbb{N}$ *on atomic arithmetical sentences,*

2. *For every* $\varphi \in \mathscr{L}_T^{\rightarrow}$, $\mathscr{E}_{Ł}^{n}(\varphi) = \mathscr{E}_{Ł}^{n}(T\ulcorner\varphi\urcorner)$,

3. $\mathscr{E}_{Ł}^{n}$ *is an n-valued Łukasiewicz evaluation function.*

Moreover, there is no total, $[0,1]$*-valued function* $\mathscr{E}_{Ł}^{\infty} : SENT_{\mathscr{L}_T^{\rightarrow}} \longmapsto [0,1]$ *s.t.:*

1. $\mathscr{E}_{Ł}^{\infty}$ *agrees with* $\mathbb{N}$ *on atomic arithmetical sentences,*

2. *For every* $\varphi \in \mathscr{L}_T^{\rightarrow}$, $\mathscr{E}_{Ł}^{\infty}(\varphi) = \mathscr{E}_{Ł}^{\infty}(T\ulcorner\varphi\urcorner)$,

3. $\mathscr{E}_{Ł}^{\infty}$ *is a* $[0,1]$*-valued Łukasiewicz evaluation function.*

The clash between Łukasiewicz evaluations and naïve truth is caused by the behavior of conditional on non-classical values. No such clash happens for strong Kleene semantics (as Kripke's theory shows). The result of Theorem 4.1.2 in the case of Łukasiewicz 3-valued semantics is sometimes used to justify a stronger claim: Łukasiewicz 3-valued semantics is *not compatible with the framework of Kripke's theory*. McGee, for example, puts this idea as follows:

> An historically important example of a method for *handling truth-value gaps* which is *not amenable to Kripke's techniques* is the 3-valued logic of Łukasiewicz [...] ([McGee 1991], p. 87, my emphasis)

[6]The proof of the first claim is a generalization of Curry's paradox, see [Field 2008], pp. 85-86. The second claim was proven in [Restall 1992]. This result is discussed in Subsection 5.4, see the proof of Theorem 5.4.3 and the relative discussion. See also [Hájek, Paris, and Shepherdson 2000].

In the following, I will show that the stronger claim is mistaken: there are ways to handle "gaps" using Łukasiewicz 3-valued semantics that are perfectly "amenable to Kripke's techniques". I will give a Kripke-style construction that uses Łukasiewicz 3-valued semantics compatibly with weak naïveté. The resulting construction, called *ŁK-construction* (for Łukasiewicz and Kripke), enjoys all the nice features of Kripke's theory and it turns out to be more expressive, thanks to the Łukasiewicz conditional: it is possible to use the ŁK-construction to validate sentences that Kripke's theory cannot consistently validate, including sentences formalizing non-trivial philosophical claims, such as (G) and (LC).[7] Finally, the conditional modeled by the ŁK-construction is simple to interpret, flexible to apply, and characterized by an genuine introduction rule, making the ŁK-construction an interesting candidate to address FIELD'S PROGRAM.

4.2 The Łukasiewicz-Kripke construction

4.2.1 Inductive definitions, gaps, and partiality

To clarify McGee's verdict that Łukasiewicz 3-valued semantics is not "amenable to Kripke's techniques" and my strategy to use it in Kripke's framework nonetheless, I will introduce a few abstract notions, adapted from [Feferman 1984]. Thanks to these notions, we see that every Łukasiewicz semantics lacks a feature which is a sufficient condition to generate fixed points of the kind that Kripke used.

Let $\mathscr{E}$ be an evaluation for $\mathscr{L}_T^{\rightarrow}$-sentences, i.e. a function $\mathscr{E} : SENT_{\mathscr{L}_T^{\rightarrow}} \longmapsto \mathbf{V}$, for a set $\mathbf{V}$ s.t. $|\mathbf{V}| \geq 2$, serving as value space.

Definition 4.2.1. (Information orderings)

A binary relation $\sqsubseteq$ on $\mathbf{V}$ is an *information ordering on* $\mathbf{V}$ if these conditions hold:

1. $\sqsubseteq$ is a partial ordering of $\mathbf{V}$,

[7] See Subsection 2.5.1.

2. there is one non-empty, proper subset of $\mathbf{V}$, call it $\mathbf{M}_{\sqsubseteq}$, whose elements are (i) all the maximal elements of $\sqsubseteq$, and (ii) s.t., if $\mathbf{v} \in \mathbf{V} \setminus \mathbf{M}_{\sqsubseteq}$, then for every $\mathbf{m} \in \mathbf{M}_{\sqsubseteq}$, $\mathbf{v} \sqsubset \mathbf{m}$ (namely $\mathbf{v} \sqsubseteq \mathbf{m}$ and $\mathbf{m} \not\sqsubseteq \mathbf{m}$).

▲

Definition 4.2.2. ($\sqsubseteq$-monotonicity)
Let $\sqsubseteq$ be an information ordering on $\mathbf{V}$ and I be an index set on $\mathbf{V}$: a function $f : \mathbf{V}^I \longmapsto \mathbf{V}$ is $\sqsubseteq$-*monotone* if, whenever $p_i \sqsubseteq q_i$ for every $i \in I$ (for $p_i, q_i \in \mathbf{V}$), then $f([p_i]_{i \in I}) \sqsubseteq f([q_i]_{i \in I})$. ▲

Following Feferman, a semantics for a language $\mathcal{L}ang$ is said to be $\sqsubseteq$-monotone if the value functions on $\mathbf{V}$ it associates with the connectives and quantifiers of $\mathcal{L}ang$ are $\sqsubseteq$-monotone. There is a basic link between $\sqsubseteq$-monotonicity and weak naïveté.

Theorem 4.2.3. (Fixed-Model Theorem (Aczel and Feferman))[8]
Let $\mathscr{L}^+$ be a language s.t. it is identical to or extends $\mathscr{L}_T^{\rightarrow}$ and it has a $\sqsubseteq$-monotone semantics for some information ordering $\sqsubseteq$. For any model $\mathscr{M}$ of PA*, there is a partial structure $(\mathscr{M}, S)$ and an evaluation function $\mathscr{E}_{(\mathscr{M},S)}$ based on it s.t. for every $\varphi \in SENT_{\mathscr{L}^+}$, $\mathscr{E}_{(\mathscr{M},S)}(T\ulcorner\varphi\urcorner) = \mathscr{E}_{(\mathscr{M},S)}(\varphi)$.*

A way to see information orderings more concretely is the following: they indicate that some values are "more definitive" than others. If $\sqsubseteq$ is an information ordering on $\mathbf{V}$, a $\sqsubseteq$-monotone value-function sends its maximal elements to maximal elements, but it sends non-maximal elements to elements possibly higher in $\sqsubseteq$. If we see value-functions as providing new information on the values of sentences,[9] the picture is that new information can upgrade non-maximal values to more definitive values, while maximal values are insensitive to new information.

[8] I adapted the Fixed-Model Theorem to my context; for a proof and a discussion, see [Feferman 1984], pp. 91-93. The Fixed-Model Theorem is a very general result that has many applications, also quite distant from truth. For the general version, see [Aczel and Feferman 1980].

[9] E.g., because they are used to go from the value of the subsentences of φ to the value of φ.

Theorem 4.2.3 applies naturally to Kripke's theory. Let $\mathbf{V}$ be $\{1,0,1/2\}$; the information ordering standardly associated with strong Kleene semantics, $\sqsubseteq_{\mathrm{K}}$, is described by the following diagram (an arrow indicates that the element at the origin of the arrow bears the relation $\sqsubseteq_{\mathrm{K}}$ to the element pointed at by the arrow):

$$\begin{tikzcd} 1 \arrow[loop left] & & 0 \arrow[loop right] \\ & 1/2 \arrow[ul] \arrow[ur] \arrow[loop below] & \end{tikzcd}$$

Figure 4.1: The information ordering $\sqsubseteq_{\mathrm{K}}$

Strong Kleene semantics is $\sqsubseteq_{\mathrm{K}}$-monotone. By Theorem 4.2.3, this is enough to extend a model of PA to a strong Kleene model for $\mathscr{L}_T^{\rightarrow}$ respecting weak naïveté.

Remark 4.2.4.

Information orderings need not have much in common with the numerical orderings often used in semantics. In strong Kleene semantics, e.g., the value of $\varphi \wedge \psi$ is the *minimum* of the values of φ and ψ with respect to the numerical ordering of $\{1,0,1/2\}$. $\sqsubseteq_{\mathrm{K}}$, instead, is related to build-up of Kripke fixed points. Let's only consider consistent fixed points. The stages of the construction of a Kripke fixed point yield partial distributions of values 1 and 0, i.e. partial evaluations, and the Kripke jump Φ is monotone on them: if two such evaluations f, g are s.t. $f \subseteq g$,[10] applying Φ to the pairs of sets generating f and g yields pairs of sets generating f' and g' s.t. $f' \subseteq g'$. $\sqsubseteq_{\mathrm{K}}$ tells us that when value 1 or 0 is assigned to a sentence in the construction of a Kripke fixed point, applying the Kripke jump one cannot change this value or go back to "not evaluated". This seems in line with Kleene's understanding of $1/2$ and justifies the name "information ordering" for $\sqsubseteq_{\mathrm{K}}$. ▲

[10]By this I mean that the function f, seen as a set, is a subset of g.

Things are much less nice for Łukasiewicz 3-valued semantics. There is no information ordering $\sqsubseteq$ with $\mathbf{V} = \{1, 0, 1/2\}$ s.t. the value-functions of Łukasiewicz 3-valued semantics are $\sqsubseteq$-monotone. The failure of $\sqsubseteq$-monotonicity is to be found in the value-function associated with the conditional, call it $f_{Ł}^{\rightarrow}$. In the case of $\sqsubseteq_{K}$, e.g., even if $1/2 \sqsubseteq_{K} 1/2$ and $1/2 \sqsubseteq_{K} 0$, we have that $f_{Ł}^{\rightarrow}(1/2, 0) \sqsubset_{K} f_{Ł}^{\rightarrow}(1/2, 1/2)$.

We can make Łukasiewicz 3-valued semantics compatible with Theorem 4.2.3, however, by *giving a different status to the non-classical value* $1/2$. So far, I identified the value $1/2$ of strong Kleene semantics with C-gaps. This identification is harmless but, as shown in Subsection 2.5.1, this is not the most Kripkean way of looking at Kripke's theory. Strictly speaking, C-gaps are not part of Kripke's theory, which does not give us a way to map *totally* $\mathscr{L}_{T}^{\rightarrow}$-sentences to 1, 0, and $1/2$, but a way to map *partially* $\mathscr{L}_{T}^{\rightarrow}$-sentences to 1 and 0.[11] If one takes 1 and 0 as the only values where to map partially $\mathscr{L}_{T}^{\rightarrow}$-sentences, however, value $1/2$ remains inevitably associated with the intractable C-gaps. But considering 1 and 0 as the only values to use in a partial evaluation is by no means necessary. *Nothing forbids us to consider* $1/2$ *as a value just like* 0 *and* 1, mapping $\mathscr{L}_{T}^{\rightarrow}$-sentences *partially* to 1, 0, *and* $1/2$. Considering $1/2$ on a par with 1 and 0 allows us to make room for another notion of gappiness, as a *positive* status for sentences and a genuine value assignment, not merely as what remains after "real" values are assigned.

Kripke's treatment of gappy sentences is conceptually unsatisfactory. Even if there may be very good reasons to avoid assigning a value to certain sentences or to assign values that have a special status to some sentences, Kripke's trichotomy (truth-sets, falsity-sets, C-gaps) is too coarse-grained. Sentences that seem very different between them are conflated together in C-gaps, but some of them may be differentiated by distinct value assignments in other inductive evaluations.[12] Moreover, even if C-gaps contain safely sentences that could generate inconsisten-

[11] For a survey of partial logics, see [Blamey 2002].

[12] E.g. the Curry-like sentences in [Field 2008], pp. 85-86. See also Subsection 5.2.3.

cies (e.g. λ), we have seen in Subsection 2.5.1 that it is impossible to use Kripke's theory to validate desirable claims based on C-gappy sentences, such as (G) ("The liar sentence is true exactly if its negation is true"). So, it seems worthwhile to explore the possibility to refine Kripke's trichotomy and to isolate those sentences (C-gappy in some Kripke fixed points) that can be evaluated in an inductive construction, using an evaluation scheme stronger than Kripke's one.

While C-gaps (of consistent Kripke fixed points) are not inductively definable,[13] a positive notion of gappiness can befit the framework of inductive definitions, and Kripke's theory can be expanded to include it. *Positive* gaps can be used together with Kripke truth- and falsity-sets to interpret the Łukasiewicz 3-valued conditional, using positive information on sentences having value 1, 0, *and now also* $1/2$. The resulting construction may have its own analog of C-gaps, i.e. sentences getting none of the 3 values. But with the positive notion of gappiness, we will be able to treat many sentences originally confined in C-gaps, making them "amenable to Kripke's techniques". I will now implement these ideas.

4.2.2 The formal construction

The first task is to turn the lacking or purely negative characterization of C-gappy sentences we have in Kripke's theory into some bit of positive information. I will now provide a suitable monotone construction that partly does the job. The construction itself does not *build* positive gaps, but rather *makes room* for a third value $1/2$. This choice has the fundamental advantage that it allows the construction to accommodate several conceptions of gappiness.[14]

Just as we can give an inductive construction that isolates the strong Kleene clauses for the preservation of values 1 and 0, we can give an inductive construction for the strong Kleene clauses for value $1/2$. Using the Kripke jump, I give an

[13] See Lemma 2.5.3.

[14] See the discussion in Subsection 4.4.1.

inductive construction that, taking a set S_3 as hypothesis for sentences valued $1/2$ and sets S_1 and S_2 for values 1 and 0, increases the former using also the latter.

Definition 4.2.5. (Positive gappiness preservation construction for $\mathscr{L}_T^{\rightarrow}$)

For any three sets $S_1, S_2, S_3 \subseteq \omega$, define the *P-gappy jump of S_3 for $\mathscr{L}_T^{\rightarrow}$ relative to S_1 and S_2* as the set S_3^* such that:

$n \in S_3^*$ if $n \in S_3$, or

(i) n is $\neg\varphi$ and $\varphi \in \mathscr{L}_T^{\rightarrow}$, and $\varphi \in S_3$; or

(ii) n is $\varphi \wedge \psi$ and $\varphi, \psi \in \mathscr{L}_T^{\rightarrow}$, and $\left\{ \begin{array}{l} \varphi \in S_3 \text{ and } \psi \in S_3, \text{ or} \\ \varphi \in S_3 \text{ and } \psi \in \mathfrak{E}_\Phi(S_1, S_2), \text{ or} \\ \varphi \in \mathfrak{E}_\Phi(S_1, S_2) \text{ and } \psi \in S_3 \end{array} \right\}$ or

(iii) n is $\varphi \vee \psi$ and $\varphi, \psi \in \mathscr{L}_T^{\rightarrow}$, and $\left\{ \begin{array}{l} \varphi \in S_3 \text{ and } \psi \in S_3, \text{ or} \\ \varphi \in S_3 \text{ and } \psi \in \mathfrak{A}_\Phi(S_1, S_2), \text{ or} \\ \varphi \in \mathfrak{A}_\Phi(S_1, S_2) \text{ and } \psi \in S_3 \end{array} \right\}$ or

(iv) n is $\forall x \chi(x)$ and $\forall x \chi(x) \in \mathscr{L}_T^{\rightarrow}$, and, for at least one $s \in CTER_{\mathscr{L}_T^{\rightarrow}}$, $\chi(s) \in S_3$, and, for all $t \neq s$, either $\chi(t) \in S_3$ or $\chi(t) \in \mathfrak{E}_\Phi(S_1, S_2)$; or

(v) n is Tt and $t = \ulcorner \chi \urcorner$ and $\chi \in \mathscr{L}_T^{\rightarrow}$ and $\chi \in S_3$.

▲

This is an inductive definition, inductive in S_1, S_2, and S_3. As usual, we associate a monotone operator with it.

Definition 4.2.6. (Positive gappiness jump)

Put $\eta(n, S_3, S_1, S_2)$ as the abbreviation of the right-hand side of "if" in Definition 4.2.5. $\eta(n, S_3, S_1, S_2)$ is a positive elementary formula, positive in S_1, S_2, and S_3. Associate with $\eta(n, S_3, S_1, S_2)$ an operator acting on triples of subsets of ω, namely

$\Psi : \wp(\omega) \times \wp(\omega) \times \wp(\omega) \longmapsto \wp(\omega) \times \wp(\omega) \times \wp(\omega)$, as follows:

$$\Psi(S_3, S_1, S_2) := \langle \{n \in \omega \mid \eta(n, S_3, S_1, S_2)\}, S_1, S_2 \rangle.$$

Ψ will be also referred to as the "P-gappy jump". ▲

Lemma 4.2.7.

The P-gappy jump is inclusive and monotone.

$\mathfrak{I}_{\Psi}(R, P, Q)$ denotes the fixed point of Ψ obtained from the triple $\langle R, P, Q \rangle$, for $P, Q, R \subseteq \omega$. Also, let $\mathfrak{H}_{\Psi}(R, P, Q)$ denote the first member of the triple $\mathfrak{I}_{\Psi}(R, P, Q)$. $\mathfrak{H}_{\Psi}(R, P, Q)$ contains the P-gappy sentences obtained from the starting hypothesis for the distribution of values $1/2$, 1, and 0 represented by $\langle R, P, Q \rangle$.

Lemma 4.2.8.

For every $P, Q, R \subseteq \omega$, $\mathfrak{H}_{\Psi}(R, P, Q)$ is inductive in P, Q, R.

The operator Ψ builds a Kripke fixed point over $\langle S_1, S_2 \rangle$, namely $\mathfrak{I}_{\Phi}(S_1, S_2)$, then it evaluates P-gappy sentences based on the sentences in S_3 and $\mathfrak{I}_{\Phi}(S_1, S_2)$ using the strong Kleene clauses, and it goes on until the fixed point.[15]

Once we have room to handle $1/2$-valued sentences, we extend Kripke's construction with clauses for $\rightarrow$ patterned after Łukasiewicz 3-valued semantics. The reading of $\rightarrow$ approximates Łukasiewicz's 3-valued conditional as the reading of $\neg$, $\wedge$, $\vee$, $\forall$ in Kripke's theory approximates their classical interpretation.[16]

[15]In giving the strong Kleene clauses for the preservation of value $1/2$, the starting hypotheses for values 1 and 0 cannot be dispensed with. Contrast this with the strong Kleene clauses for preserving value 1 (but the same applies to value 0): one can give all the conditions under which a sentence is given value 1 (0) in a strong Kleene evaluation using only clauses about value 1 (0) in that evaluation (see [Halbach 2011], pp. 202 and following). Giving all the conditions under which a sentence gets value $1/2$ in a strong Kleene evaluation, instead, requires to use all three values. For example, not all the conditions under which $\varphi \wedge \psi$ has value $1/2$ in a strong Kleene evaluation can be given using only its subsentences or negated subsentences having value $1/2$.

[16]This answers Martin's question "of what is the new conditional a generalization" for this construction (see Subsection 3.5.4). See also Subsection 3.6.2 for a discussion on approximations.

Definition 4.2.9. (ŁK-construction for $\mathscr{L}_T^{\rightarrow}$)

For any three sets $S_1, S_2, S_3 \subseteq \omega$, define as their simultaneous *ŁK-jump* the sets S_1^{T}, S_2^{F}, S_3^{G} that satisfy the following conditions:

1. $n \in S_1^{\mathrm{T}}$ if:

   (i) $n \in \mathfrak{E}_\Phi(S_1, S_2)$, or

   (ii) n is $\varphi \rightarrow \psi$ and φ, $\psi \in \mathscr{L}_T^{\rightarrow}$, and $\begin{cases} \varphi \in \mathfrak{A}_\Phi(S_1, S_2), \text{ or} \\ \psi \in \mathfrak{E}_\Phi(S_1, S_2), \text{ or} \\ \varphi, \psi \in \mathfrak{H}_\Psi(S_3, S_1, S_2). \end{cases}$

2. $n \in S_2^{\mathrm{F}}$ if:

   (i) $n \in \mathfrak{A}_\Phi(S_1, S_2)$, or

   (ii) n is $\varphi \rightarrow \psi$ and φ, $\psi \in \mathscr{L}_T^{\rightarrow}$, and $\varphi \in \mathfrak{E}_\Phi(S_1, S_2)$, $\psi \in \mathfrak{A}_\Phi(S_1, S_2)$.

3. $n \in S_3^{\mathrm{G}}$ if:

   (i) $n \in \mathfrak{H}_\Psi(S_3, S_1, S_2)$, or

   (ii) n is $\varphi \rightarrow \psi$ and φ, $\psi \in \mathscr{L}_T^{\rightarrow}$, and

$$\begin{cases} \varphi \in \mathfrak{E}_\Phi(S_1, S_2) \text{ and } \psi \in \mathfrak{H}_\Psi(S_3, S_1, S_2), \text{ or} \\ \varphi \in \mathfrak{H}_\Psi(S_3, S_1, S_2) \text{ and } \psi \in \mathfrak{A}_\Phi(S_1, S_2). \end{cases}$$

▲

S_1^{T}, S_2^{F}, S_3^{G} are inductive in S_1, S_2, S_3. Let's define the associated operator.

Definition 4.2.10.

Let $\vartheta_1(n, S_1, S_2, S_3)$, $\vartheta_2(n, S_1, S_2S_3)$, and $\vartheta_3(n, S_1, S_2S_3)$ abbreviate the right-hand side of items 1, 2 and 3 of Definition 4.2.9, respectively. Associate with these three positive elementary formulae, positive in S_1, S_2 and S_3, a monotone operator on triples of sets. Define $\Upsilon : \wp(\omega)\times\wp(\omega)\times\wp(\omega) \longmapsto \wp(\omega)\times\wp(\omega)\times\wp(\omega)$ as:

$$\Upsilon(S_1,S_2,S_3) := \langle \{n\in\omega \mid \vartheta_1(n,S_1,S_2,S_3)\}, \{n\in\omega \mid \vartheta_2(n,S_1,S_2,S_3)\}, \\ \{n\in\omega \mid \vartheta_3(n,S_1,S_2,S_3)\}\rangle.$$

▲

$\mathfrak{I}_\Upsilon(P,Q,R)$ is the fixed point of Υ obtained from $\langle P,Q,R\rangle$ and is inductive in P, Q, and R. By a small notational abuse, I will use P_∞ to denote the first component of $\mathfrak{I}_\Upsilon(P,Q,R)$ (similarly for Q_∞ or R_∞).[17]

Let's say that a fixed point $\mathfrak{I}_\Upsilon(P,Q,R)$ is *consistent* if P_∞, Q_∞, and R_∞ are pairwise disjoint and *inconsistent* otherwise. Inconsistent fixed points of the ŁK-construction are comparable in role to inconsistent fixed points of Kripke's construction, being useful to indicate some expressive limitations of the theory.[18]

Finally, as with Kripke's construction, we can associate an evaluation with Υ.

Definition 4.2.11. (ŁK-evaluation)

For every $P,Q,R\subseteq\omega$, define the *ŁK evaluation generated by* $\mathfrak{I}_\Upsilon(P,Q,R)$ as the set $\mathscr{P}_{PQR}\subseteq SENT_{\mathscr{L}_T}\times\{1,0,1/2\}$ such that:

$$\langle\varphi,\mathbf{v}\rangle\in\mathscr{P}_{PQR} \text{ iff } \begin{cases} \varphi\in P_\infty \text{ and } \mathbf{v} \text{ is } 1, \text{ or} \\ \varphi\in Q_\infty \text{ and } \mathbf{v} \text{ is } 0, \text{ or} \\ \varphi\in R_\infty \text{ and } \mathbf{v} \text{ is } 1/2 \end{cases}$$

Clearly, $\mathscr{P}_{PQR}$ is a *function* if and only if $\mathfrak{I}_\Upsilon(P,Q,R)$ is consistent. For a consistent fixed point $\mathfrak{I}_\Upsilon(P,Q,R)$, I will simply write $\mathscr{P}_{PQR}(\varphi)=\mathbf{v}$. ▲

The jump Υ takes as argument a triple $\langle P,Q,R\rangle$ of subsets of ω, reading it as the starting distribution of values 1, 0, and 1/2 *in this order*; then it builds Kripke truth- and falsity-sets over P and Q and fixed points of the P-gappy construction over R, P,

[17] Strictly speaking, writing P_∞ alone is not meaningful, as to know what is in it we must know the other members of the triple, but when writing P_∞, Q_∞, or R_∞ alone, the triple will always be clear.

[18] See Section 2.3.

and Q; finally Υ evaluates the conditionals $\varphi \to \psi$ s.t. φ, or ψ, or both are in these sets, according to the Łukasiewicz clauses. The resulting new triple $\langle P^{\mathrm{T}}, Q^{\mathrm{F}}, R^{\mathrm{G}} \rangle$ preserves the initial triple and possibly increases it. For example, P^{T} includes P and adds to it the Kripke truth-set over P and Q, plus those conditionals that are given value 1 according to the Łukasiewicz clauses, using the starting distribution. More specifically, a conditional $\varphi \to \psi$ is in P^{T} in three cases:

- its antecedent φ is in the Kripke falsity-set built over $\langle P, Q \rangle$, our starting hypotheses for the distribution of values 1 and 0, part of the overall starting hypothesis $\langle P, Q, R \rangle$ for the distribution of 1, 0, and 1/2; or
- its consequent ψ is in the Kripke truth-set built over $\langle P, Q \rangle$; or
- both its antecedent φ and its consequent ψ are in the set of P-gappy sentences built over the same distribution of values.

The content of Q^{F} and R^{G} is explained informally in a similar way. Once the new triple $\langle P^{\mathrm{T}}, Q^{\mathrm{F}}, R^{\mathrm{G}} \rangle$ is constructed, it is used to build new Kripke truth- and falsity-sets and fixed points of the P-gappy construction, before evaluating another round of conditionals. The construction keeps "zigzagging" in this way between the evaluations of non-conditionals and conditionals, until it reaches a fixed point, where it is closed under the Łukasiewicz clauses for preservation of values 1, 0, and 1/2 for connectives and quantifiers, plus the Kripkean clause for naïve truth.

4.2.3 Paradigms of partiality

As Kripke's theory, the ŁK-construction interprets $\mathscr{L}_T^{\to}$-sentences via fixed points, but now we have triples rather than pairs. The resulting evaluations provide partial, three-valued semantics, which we can extend to total ones, assigning an extra value, call it **N**, to sentences that do not get value 1, 0, or 1/2.

Definition 4.2.12. (C-gappy ŁK-evaluation)

For every $P,Q,R \subseteq \omega$, define the *C-gappy ŁK evaluation generated by* $\mathfrak{I}_\Upsilon(P,Q,R)$ as the set $\mathscr{P}^+_{PQR} \subseteq SENT_{\mathscr{L}_T} \times \{\mathit{1},\mathit{0},{}^1\!/\!_2,\mathbf{N}\}$ such that:

$$\langle\varphi,\mathbf{v}\rangle \in \mathscr{P}^+_{PQR} \text{ iff } \begin{cases} \varphi \in P_\infty \text{ and } \mathbf{v} \text{ is } 1, \text{ or} \\ \varphi \in Q_\infty \text{ and } \mathbf{v} \text{ is } 0, \text{ or} \\ \varphi \in R_\infty \text{ and } \mathbf{v} \text{ is } 1/2, \text{ or} \\ \varphi \in SENT_{\mathscr{L}_T^{\rightarrow}} \setminus (P_\infty \cup Q_\infty \cup R_\infty) \text{ and } \mathbf{v} \text{ is } \mathbf{N} \end{cases}$$

Just as $\mathscr{P}_{PQR}$, also $\mathscr{P}^+_{PQR}$ is a *function* whenever $\mathfrak{I}_\Upsilon(P,Q,R)$ is consistent. ▲

C-gappy ŁK-evaluations include strictly Łukasiewicz 3-valued evaluations and are $\sqsubseteq$-monotone on the following information ordering:

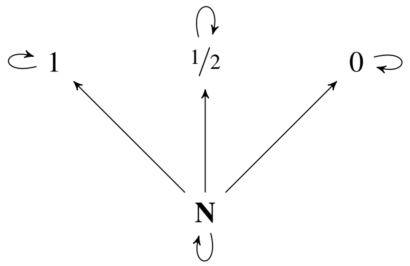
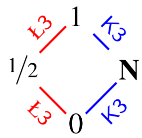


Figure 4.2: The information ordering $\sqsubseteq_{\text{ŁK}}$

This shows, in the light of Theorem 4.2.3, how it is possible to get evaluations that respect weak naïveté for the evaluation scheme I employ. Hartry Field suggested a nice way of presenting the interpretation of the conditional supported by the ŁK-construction. Since the value $\mathbf{N}$ is not strictly speaking part of the ŁK-construction, it represents a lack of value, and it behaves just as value $1/2$ of strong Kleene semantics, one could see the ŁK-construction as showing that it is possible to keep Łukasiewicz 3-valued semantics and strong Kleene semantics together in the reading of the conditional: in every consistent fixed point of Υ, it is possible to give the stronger Łukasiewicz reading to some conditionals, where the remaining ones could be understood as strong Kleene conditionals. The evaluation relations between the (now) 4 values can be summed up in the following “mixed" Hasse diagram, and they are specified by the labels on its lines:



If $f_{ŁK}^{\rightarrow}$ is the value-function associated with $\rightarrow$ by the ŁK-construction, $f_{ŁK}^{\rightarrow}(1/2, 1/2) = 1$ but $f_{ŁK}^{\rightarrow}(\mathbf{N},\mathbf{N}) = \mathbf{N}$. $f_{ŁK}^{\rightarrow}$ behaves as the value-function of the Łukasiewicz 3-valued conditional for all inputs in which a value other than **N** results, and as the corresponding strong Kleene value-function in the remaining cases.

However, probably the best way to look at my strategy to make Łukasiewicz 3-valued semantics compatible with Kripke's framework is not to see consistent fixed points of Υ as giving a new 4-valued semantics. My discussion on the status of C-gappy sentences in Kripke's theory[19] suggests that what is really crucial in understanding constructions *à la* Kripke is not the specific semantics which is used, but rather the *paradigm of partiality which is adopted*. Let me explain this idea with an example. Take Kleene's strong and weak semantics. They can be seen as two semantics in their own right which include classical value-functions, or as two ways for classical semantics to be partial, i.e. to fail to assign a value to some sentence. Weak and strong Kleene semantics are $\sqsubseteq_{\mathrm{K}}$-monotone but they diverge in a key aspect: in the strong case, the fact that only some sub-sentence ψ of a sentence φ is given a $\sqsubseteq_{\mathrm{K}}$-maximal value may suffice to give a $\sqsubseteq_{\mathrm{K}}$-maximal value to the whole φ, whereas in the weak case this is not so. In strong Kleene semantics, e.g., a disjunction with one 1-valued disjunct and one 1/2-valued disjunct gets value 1, but this is not so in the weak variant. This key aspect tells strong and

[19] See Subsection 2.5.1.

weak Kleene semantics apart, but it can be isolated and generalized beyond the values and information orderings used in them: we can apply the strong or the weak Kleene way of making a semantics partial *to a non-classical semantics as well*. Supervaluationism may be seen as a partiality paradigm in my sense too, a way to render the value-functions used in a certain semantics partial. Supervaluationist semantics validate all the classical logical truths, but one could adapt the supervaluationist machinery to recover the Łukasiewicz 3-valued truths, or others. This way of conceiving strong Kleene, weak Kleene, and supervaluationist semantics may explain Kripke's claim that he did not adopt a non-classical semantics: he seems to see his constructions as imposing partiality paradigms on classical semantics.[20]

In the ŁK-construction, I make Łukasiewicz 3-valued semantics partial via the strong Kleene paradigm. The similarity between the strong Kleene partial version of classical semantics and the strong Kleene partial version of Łukasiewicz 3-valued semantics is clear from Figures 4.1 and 4.2. The choice of the *strong* Kleene paradigm is evident from Definitions 4.2.5 and 4.2.9.: e.g., if a conditional has a 1-valued consequent and a **N**-valued antecedent (or a valueless antecedent, if one does not add the extra value), the whole conditional has value 1. Adopting the weak Kleene paradigm, the whole conditional would have value **N** (or no value).

The point I try to make here is not merely terminological. Rather than identifying always new semantics generated by new values and value-functions, identifying the relevant paradigm of partiality seems the most telling and Kripkean way to understand constructions *à la* Kripke. Moreover, focusing on partiality paradigms is conceptually fruitful, since it allows to push forward some frameworks and tools to develop new theories, as it was done with the ŁK-construction. In Chapter 5, we will see how this perspective can open the way to much more expressive theories, that include resources that go beyond the Kripkean setting.

[20]See footnote 22 in Subsection 2.5.1.

4.3 Properties of the ŁK-construction

I will now review the main facts about the ŁK-construction. In the first place, the truth predicate interpreted by the ŁK-construction enjoys the same forms of naïveté that it enjoys in Kripke's theory, namely weak naïveté and intersubstitutivity of truth. Here, however, gappiness enters the picture explicitly via P-gaps.

Theorem 4.3.1. (Weak Naïveté)

For every $\varphi \in \mathscr{L}_T^{\rightarrow}$ and every $P, Q, R \subseteq \omega$:

1. $\varphi \in P_\infty$ *(Q_∞, R_∞) iff* $T\ulcorner\varphi\urcorner \in P_\infty$ *(Q_∞, R_∞).*

2. $\varphi \in SENT_{\mathscr{L}_T^{\rightarrow}} \setminus (P_\infty \cup Q_\infty \cup R_\infty)$ *iff* $T\ulcorner\varphi\urcorner \in SENT_{\mathscr{L}_T^{\rightarrow}} \setminus (P_\infty \cup Q_\infty \cup R_\infty)$.

Proof.

Ad 1, observe that:

(4.1)

$$\begin{aligned} T\ulcorner\varphi\urcorner \in P_\infty \quad &\text{iff} \quad T\ulcorner\varphi\urcorner \in \mathfrak{E}_\Phi(P_\infty, Q_\infty) \quad \text{(by def. and fixed-point property of } \Upsilon) \\ &\text{iff} \quad \varphi \in \mathfrak{E}_\Phi(P_\infty, Q_\infty) \qquad \text{(by Theorem 2.4.1)} \end{aligned}$$

Replacing P_∞ with Q_∞ (R_∞) and $\mathfrak{E}_\Phi(P_\infty, Q_\infty)$ with $\mathfrak{A}_\Phi(P_\infty, Q_\infty)$ ($\mathfrak{H}_\Psi(R_\infty, P_\infty, Q_\infty)$ respectively) in equivalence (4.1) completes the proof of item 1.

Ad 2, let $\varphi \in SENT_{\mathscr{L}_T^{\rightarrow}} \setminus (P_\infty \cup Q_\infty \cup R_\infty)$. If $T\ulcorner\varphi\urcorner \in P_\infty$ (or Q_∞ or R_∞), by equivalence (4.1) also $\varphi \in P_\infty$, so $\varphi \in P_\infty \cap SENT_{\mathscr{L}_T^{\rightarrow}} \setminus (P_\infty \cup Q_\infty \cup R_\infty)$, which is absurd. The same holds starting with $T\ulcorner\varphi\urcorner$. □

Corollary 4.3.2. (Intersubstitutivity of Truth)[21]

For every $\varphi, \psi, \chi \in \mathscr{L}_T^{\rightarrow}$ and $P, Q, R \subseteq \omega$, if ψ and χ are alike except that ψ has an occurrence of φ of $\mathscr{L}_T$ where χ has an occurrence of $T\ulcorner\varphi\urcorner$, then:

$$\psi \in P_\infty\ (Q_\infty,\ R_\infty) \quad \textit{iff} \quad \chi \in P_\infty\ (Q_\infty,\ R_\infty,\ \textit{respectively}).$$

[21] By Theorem 4.3.1, the proof is as in Corollary 2.4.2.

The main interest of these results is that, together with the facts reviewed in Subsection 4.2.2, they disprove McGee's verdict, giving a way to interpret Łukasiewicz 3-valued semantics that is "amenable to Kripke's techniques". Indeed, only "Kripke's techniques" were used here: as in Kripke's theory, we have an inductive definition (the ŁK-construction) and a monotone jump inductive in its arguments (Υ). The clauses to interpret T are Kripke's ones: for every set inductively built in the construction, if φ is in that set, so is $T\ulcorner\varphi\urcorner$. Łukasiewicz 3-valued semantics, thus, has a use that makes it compatible with Kripke's framework.

The main features of my way to use Kripke's techniques on Łukasiewicz 3-valued semantics are positive gaps and partiality. On partiality, if a fixed point of Υ is consistent, some sentence receives no value in it – as in Kripke's theory.[22]

Lemma 4.3.3.
If $\Im_\Upsilon(P,Q,R)$ is s.t. for every $\varphi \in \mathscr{L}_T^{\rightarrow}$, $\varphi \in (P_\infty \cup Q_\infty \cup R_\infty)$, then $\Im_\Upsilon(P,Q,R)$ is inconsistent.

Proof.
Let ι be $Tt_\iota \rightarrow \neg Tt_\iota$ for a term t_ι s.t. $t_\iota = \ulcorner Tt_\iota \rightarrow \neg Tt_\iota\urcorner$. If $\iota \in P_\infty$, by the fixed-point property, $T\ulcorner\iota\urcorner \in P_\infty$, $\neg T\ulcorner\iota\urcorner \in Q_\infty$, and $T\ulcorner\iota\urcorner \rightarrow \neg T\ulcorner\iota\urcorner \in Q_\infty$, i.e. $\iota \in Q_\infty$, which means that $\Im_\Upsilon(P,Q,R)$ is inconsistent. The same conclusion follows from the supposition that $\iota \in Q_\infty$ or that $\iota \in R_\infty$. $\square$

So, every consistent fixed point $\Im_\Upsilon(P,Q,R)$ has its own kind of C-gaps, i.e. $SENT_{\mathscr{L}_T^{\rightarrow}} \setminus (P_\infty \cup Q_\infty \cup R_\infty)$. This result entails that, just as Kripke's theory, the ŁK-construction has no laws:[23] there is no schematic law s.t. all its instances have value 1 in at least one consistent fixed point of Υ. Once this price is paid, however, the ŁK-construction improves on Kripke's semantics for the logical vocabulary and thus on the treatment of truth-theoretic facts, as the following results indicate.

[22] Compare Lemma 2.5.1.
[23] Recall that I consider only the strong Kleene version of Kripke's theory.

Corollary 4.3.4. (Strong(er) Naïveté)

For every $\varphi \in \mathscr{L}_T^{\rightarrow}$ *and* $P,Q,R \subseteq \omega$*, if* $\varphi \in (P_\infty \cup Q_\infty \cup R_\infty)$*,* $\varphi \leftrightarrow T\ulcorner\varphi\urcorner \in P_\infty$*.*

Proof.

Let $\varphi \in P_\infty$. By Theorem 4.3.1, $T\ulcorner\varphi\urcorner \in P_\infty$. Then, both $\varphi \rightarrow T\ulcorner\varphi\urcorner \in (P_\infty)^{\mathrm{T}}$ and $T\ulcorner\varphi\urcorner \rightarrow \varphi \in (P_\infty)^{\mathrm{T}}$. By the fixed-point property, $P_\infty = (P_\infty)^{\mathrm{T}}$. So $\varphi \leftrightarrow T\ulcorner\varphi\urcorner \in \mathfrak{E}_\Phi(P_\infty, Q_\infty) \subseteq (P_\infty)^{\mathrm{T}} = P_\infty$. If $\varphi \in Q_\infty$ or $\varphi \in R_\infty$, the case is dual. $\square$

Corollary 4.3.5. (Restricted Compositionality)

For every $\varphi \in \mathscr{L}_T^{\rightarrow}$ *and* $P,Q,R \subseteq \omega$*, if* $\varphi \in (P_\infty \cup Q_\infty \cup R_\infty)$*, then:*

$(T\neg)$ *If* φ *is* $\neg\psi$*, then* $(T\ulcorner\neg\psi\urcorner \leftrightarrow \neg T\ulcorner\psi\urcorner) \in P_\infty$*.*

$(T\wedge)$ *If* φ *is* $\psi \wedge \chi$*, then* $(T\ulcorner\psi \wedge \chi\urcorner \leftrightarrow (T\ulcorner\psi\urcorner) \wedge (T\ulcorner\chi\urcorner)) \in P_\infty$*.*

$(T\vee)$ *If* φ *is* $\psi \vee \chi$*, then* $(T\ulcorner\psi \vee \chi\urcorner \leftrightarrow (T\ulcorner\psi\urcorner) \vee (T\ulcorner\chi\urcorner)) \in P_\infty$*.*

$(T\rightarrow)$ *If* φ *is* $\psi \rightarrow \chi$*, then* $(T\ulcorner\psi \rightarrow \chi\urcorner \leftrightarrow (T\ulcorner\psi\urcorner) \rightarrow (T\ulcorner\chi\urcorner)) \in P_\infty$*.*

$(T\forall)$ *If* φ *is* $\forall x\psi(x)$*, then* $(T\ulcorner\forall x\psi(x)\urcorner \leftrightarrow \forall x T\ulcorner\psi(\dot{x})\urcorner) \in P_\infty$*.*

Proof.

If $\varphi \in (P_\infty \cup Q_\infty \cup R_\infty)$, then $\varphi \leftrightarrow \varphi \in P_\infty$, so the result holds by Corollary 4.3.2. $\square$

Lemma 4.3.6. (Fieldian Determinateness)

For every $\varphi \in \mathscr{L}_T^{\rightarrow}$ *and* $P,Q,R \subseteq \omega$*, the following holds:*

1. *Let* $\mathcal{D}(\varphi)$ *stand for* $\neg(\varphi \rightarrow \neg\varphi)$*. If* $\varphi \in Q_\infty \cup R_\infty$*, then* $\neg\mathcal{D}(\varphi) \in P_\infty$*.*

2. *If* $\mathfrak{I}_\Upsilon(P,Q,R)$ *is consistent and* $\varphi \leftrightarrow \neg\varphi \in P_\infty$*, then* $\varphi, \neg\varphi \in R_\infty$

Proof.

Ad 1, let $\varphi \in R_\infty$. By the fixed-point property of Υ, $\neg\varphi \in R_\infty$, $\varphi \rightarrow \neg\varphi \in P_\infty$, $\neg\neg(\varphi \rightarrow \neg\varphi) \in P_\infty$, i.e. $\neg\mathcal{D}(\varphi) \in P_\infty$. The same holds letting $\varphi \in Q_\infty$.

Ad 2, let $\mathfrak{I}_\Upsilon(P,Q,R)$ be consistent and $\varphi \leftrightarrow \neg\varphi \in P_\infty$. By the fixed-point property of P_∞, both the following hold:

(i) either $(\varphi \in Q_\infty)$, or $(\neg\varphi \in P_\infty)$, or $(\varphi, \neg\varphi \in R_\infty)$;

(ii) either $(\neg\varphi \in Q_\infty)$, or $(\varphi \in P_\infty)$, or $(\neg\varphi, \varphi \in R_\infty)$.

If $\varphi \in Q_\infty$, then $\varphi \notin P_\infty$, $\varphi \notin R_\infty$, and $\neg\varphi \notin Q_\infty$, so item (ii) cannot hold if $\varphi \in Q_\infty$ (the same follows if $\neg\varphi \in P_\infty$). Dually, if $\neg\varphi \in Q_\infty$ or $\varphi \in P_\infty$, condition (i) cannot obtain. Since $\mathfrak{I}_\Upsilon(P,Q,R)$ is consistent, $\varphi \in R_\infty$ and $\neg\varphi \in R_\infty$. $\square$

Corollaries 4.3.4, 4.3.5 and Lemma 4.3.6 show how the ŁK-construction improves on Kripke's construction in enabling us to validate claims about the semantics of $\mathscr{L}_T^{\rightarrow}$, formalized in the object-language itself. In Kripke's theory, Tarski $\equiv$-biconditionals hold only for elements of Kripke truth- and falsity-sets and not for their C-gaps. Despite the fact that there are valueless sentences in consistent fixed points of Υ as well (so, an analogue of C-gaps for the ŁK-construction), here we also have P-gaps at our disposal, and P-gappy sentences are quite tractable. Corollary 4.3.4 tells us that we can validate Tarski $\leftrightarrow$-biconditionals for P-gappy sentences as well. A similar remark goes for Corollary 4.3.2 and intersubstitutivity of truth, and for Corollary 4.3.5 and the compositional behavior of naïve truth.

Lemma 4.3.6 shows that the ŁK-construction enables us to define a Fieldian determinateness operator, which declares in the object-language that, for every fixed point $\mathfrak{I}_\Upsilon(P,Q,R)$, every sentence in Q_∞ or in R_∞ is "not determinately true" in $\mathfrak{I}_\Upsilon(P,Q,R)$ – by Corollary 4.3.2, "not determinate" and "not determinately true" are interchangeable. The operator $\mathcal{D}$ can be applied consistently to standard paradoxes, such as the liar sentence and Curry's sentence, asserting positively their lack of determinate truth. For example, the fixed point $\mathfrak{I}_\Upsilon(\emptyset,\emptyset,\{\lambda\})$ is consistent, and $\mathscr{P}_{\emptyset\emptyset\{\lambda\}}(\neg\mathcal{D}(\lambda)) = 1$ – of course, also $\mathscr{P}_{\emptyset\emptyset\{\lambda\}}(\lambda \leftrightarrow \neg\lambda) = 1$, validating a formal version of statement (G) from Subsection 2.5.1. No device comparable to $\mathcal{D}$ is available in Kripke's theory: since C-gaps are not usable in Kripke fixed points to validate or refute any claim, the only sentences that Kripke's theory can deem "not determinate" are the sentences in the falsity-set of a fixed point, whose lack of truth can already be captured via the strong Kleene negation. Clearly, the notion of determinateness afforded by the ŁK-construction is quite weak, since it cannot treat a revenge paradox involving it: e.g., no consistent fixed point of Υ gives a value to the sentence $\lambda^\star$ equivalent to $\neg\mathcal{D}T\ulcorner\lambda^\star\urcorner$. Despite their simplicity, however, fixed points of Υ are expressive enough to make positive claims about the indeterminateness of their P-gappy sentences. Moreover, consistent fixed points of Υ do not give to any sentence an unintended value: if $\langle P_\infty, Q_\infty, R_\infty\rangle$ is consistent and $\varphi \in Q_\infty \cup R_\infty$, then $\neg\mathcal{D}(\varphi) \in P_\infty$ and $\neg\mathcal{D}(\varphi) \notin (Q_\infty \cup R_\infty)$. In consistent fixed points, revenge paradoxes get no semantic value. Contrast this with Field's theory, a total theory, where every sentence has a value, and sometimes an unintended one.[24] The ŁK-construction, although less expressive than Field's theory, makes the notion of determinateness more uniform, as seen with Lemma 4.3.6.

The real interest of the results of this section depends on finding interesting consistent fixed points of Υ. We have already seen the case of $\mathfrak{I}_\Upsilon(\emptyset,\emptyset,\{\lambda\})$, but we would like to have interesting *families* of consistent fixed points of Υ: I will present two such cases in Section 4.5. Before that, however, I will complete an analysis of the general features of the ŁK-construction, which apply to every fixed point of Υ, thus also to the applications presented at the end of the chapter.

4.4 The ŁK-construction and FIELD'S PROGRAM

The ŁK-construction addresses FIELD'S PROGRAM nicely, as it preserves Kripke's theory and adds to it a well-behaved conditional (according to **(W1)**-**(W3)**).[25]

[24] See the results recalled in Subsection 3.6.1.
[25] See Section 2.7 or footnote 1 of Section 4.1.

4.4.1 Recovering Kripke's theory

The ŁK-construction preserves Kripke's theory in a strong sense: for every Kripke fixed point $\mathfrak{I}_\Phi(P,Q)$, there is a fixed point of Υ that is substantially identical to it, in a way that will be made precise in the following.

As I mentioned, the P-gappy jump Ψ does not *build* sets of P-gappy sentences. While Φ assigns values 1 and 0 based on some starting sets *and* arithmetical atomic truths and falsities, the P-gappy jump does not give value $1/2$ to any atomic arithmetical sentence.[26] As a result, the extension of a P-gap depends only on the starting hypotheses. This feature was designed to allow the ŁK-construction to apply to sets embodying various conceptions of gappiness. The following lemma shows that the ŁK-construction has room for literally *every* notion of gappiness, as an empty initial choice generates no P-gap.

Lemma 4.4.1. [27]

For every $P,Q \subseteq \omega$, $\mathfrak{H}_\Psi(\emptyset,P,Q) = \emptyset$.

A corollary is straightforward.

Corollary 4.4.2.

For $P,Q \subseteq \omega$, let $\langle P^\emptyset_\alpha, Q^\emptyset_\alpha, R^\emptyset_\alpha\rangle := \Upsilon^\alpha(P,Q,\emptyset)$. For all ordinals and all $\varphi,\psi \in \mathscr{L}_T^{\rightarrow}$:

1. $\varphi \rightarrow \psi \in P^\emptyset_\alpha$ *iff either* $\varphi \rightarrow \psi \in P$ *or* $\neg\varphi \vee \psi \in P^\emptyset_\alpha$
2. $\varphi \rightarrow \psi \in Q^\emptyset_\alpha$ *iff either* $\varphi \rightarrow \psi \in Q$ *or* $\varphi \wedge \neg\psi \in P^\emptyset_\alpha$
3. $R^\emptyset_\alpha = \emptyset$

Proof.

I will do the case where $P = \emptyset = Q$, but the general case is not substantially different. The proof is by induction on the build-up of $\mathfrak{I}_\Upsilon$. The base case is as follows.

[26] I assume that, for every well-established notion of gappiness, there are no gappy arithmetical sentences: this may be not entirely uncontroversial, but discussing it would be out of place here.

[27] A trivial induction establishes the result.

$$\mathfrak{I}^1_\Upsilon = \begin{array}{l} \left\langle \mathfrak{E}_\Phi \cup \{\varphi \to \psi \in \mathscr{L}_T^{\to} | \begin{cases} \varphi \in \mathfrak{A}_\Phi, \text{ or} \\ \psi \in \mathfrak{E}_\Phi, \text{ or} \\ \varphi, \psi \in \mathfrak{H}_\Psi \end{cases} \}; \right. \\ \mathfrak{A}_\Phi \cup \{\varphi \to \psi \in \mathscr{L}_T^{\to} | \varphi \in \mathfrak{E}_\Phi, \psi \in \mathfrak{A}_\Phi\}; \\ \left. \mathfrak{H}_\Psi \cup \{\varphi \to \psi \in \mathscr{L}_T^{\to} | \begin{cases} \varphi \in \mathfrak{E}_\Phi, \psi \in \mathfrak{H}_\Psi, \text{ or} \\ \varphi \in \mathfrak{H}_\Psi, \psi \in \mathfrak{A}_\Phi \end{cases} \} \right\rangle \end{array}$$

$$= \begin{array}{l} \left\langle \mathfrak{E}_\Phi \cup \{\varphi \to \psi \in \mathscr{L}_T^{\to} | \neg\varphi \vee \psi \in \mathfrak{E}_\Phi\}; \right. \\ \left. \mathfrak{A}_\Phi \cup \{\varphi \to \psi \in \mathscr{L}_T^{\to} | \varphi \wedge \neg\psi \in \mathfrak{E}_\Phi\}; \emptyset \right\rangle \end{array}$$

Let the claim hold for all ordinals $\beta \leq \alpha$ by Inductive Hypothesis.

$$\mathfrak{I}^{\alpha+1}_\Upsilon = \begin{array}{l} \left\langle \mathfrak{E}_\Phi(P^\emptyset_\alpha, Q^\emptyset_\alpha) \cup \{\varphi \to \psi \in \mathscr{L}_T^{\to} | \begin{cases} \varphi \in \mathfrak{A}_\Phi(P^\emptyset_\alpha, Q^\emptyset_\alpha), \text{ or} \\ \psi \in \mathfrak{E}_\Phi(P^\emptyset_\alpha, Q^\emptyset_\alpha), \text{ or} \\ \varphi, \psi \in \mathfrak{H}_\Psi(\emptyset, P^\emptyset_\alpha, Q^\emptyset_\alpha) \end{cases} \}; \right. \\ \mathfrak{A}_\Phi(P^\emptyset_\alpha, Q^\emptyset_\alpha) \cup \{\varphi \to \psi \in \mathscr{L}_T^{\to} | \varphi \in \mathfrak{E}_\Phi(P^\emptyset_\alpha, Q^\emptyset_\alpha), \psi \in \mathfrak{A}_\Phi(P^\emptyset_\alpha, Q^\emptyset_\alpha)\}; \\ \mathfrak{H}_\Psi(\emptyset, P^\emptyset_\alpha, Q^\emptyset_\alpha) \cup \\ \left. \{\varphi \to \psi \in \mathscr{L}_T^{\to} | \begin{cases} \varphi \in \mathfrak{E}_\Phi(P^\emptyset_\alpha, Q^\emptyset_\alpha), \psi \in \mathfrak{H}_\Psi(\emptyset, P^\emptyset_\alpha, Q^\emptyset_\alpha), \text{ or} \\ \varphi \in \mathfrak{H}_\Psi(\emptyset, P^\emptyset_\alpha, Q^\emptyset_\alpha), \psi \in \mathfrak{A}_\Phi(P^\emptyset_\alpha, Q^\emptyset_\alpha) \end{cases} \} \right\rangle \end{array}$$

$$= \begin{array}{l} \left\langle \mathfrak{E}_\Phi(P^\emptyset_\alpha, Q^\emptyset_\alpha) \cup \{\varphi \to \psi \in \mathscr{L}_T^{\to} | \neg\varphi \vee \psi \in \mathfrak{E}_\Phi(P^\emptyset_\alpha, Q^\emptyset_\alpha); \right. \\ \left. \mathfrak{A}_\Phi(P^\emptyset_\alpha, Q^\emptyset_\alpha) \cup \{\varphi \to \psi \in \mathscr{L}_T^{\to} | \varphi \wedge \neg\psi \in \mathfrak{E}_\Phi(P^\emptyset_\alpha, Q^\emptyset_\alpha)\}; \emptyset \right\rangle \end{array}$$

The limit case is straightforward from the successor one, so I will skip it. All the identities of the above proof hold by Lemma 4.4.1, the IH, and by the properties of $\mathfrak{E}_\Phi$ and $\mathfrak{A}_\Phi$. □

This corollary shows that the ŁK-construction is a generalization of Kripke's theory, since it can define also all Kripke fixed points, not adding anything irrelevant to them (i.e. reading $\rightarrow$ as a notational variant of $\supset$ and adding $\emptyset$ as P-gap). So, the fixed points of Υ range from essentially Kripke fixed points to consistent fixed points validating claims that Kripke's theory cannot account for in any sense, such as statement (G), or $\lambda \leftrightarrow \kappa$. These sentences do not belong to every fixed point of Υ, as this would impose some fixed features on the notion of gappiness and would make it impossible to recover Kripke fixed points in the above sense.

This fact draws our attention to *non-minimal* fixed points of Υ. Differently from Kripke's theory, whose $\leq_S$-least fixed point $\mathfrak{I}_\Phi$ is often considered to be the most interesting one,[28] the $\leq_S$-least fixed point of Υ is as interesting as $\mathfrak{I}_\Phi$. Non-minimal fixed points of Υ are then more than a mere technical possibility, as they can be used to model various intuitions about gappiness. The existence of "rich" non-minimal consistent fixed points of Υ follows from standard considerations.

Lemma 4.4.3. (Maximally consistent fixed points of Υ)[29]
If $\mathfrak{I}_\Upsilon(P,Q,R)$ is consistent, then there exist at least three sets $P^M, Q^M, R^M \subseteq \omega$ s.t.: $P \subseteq P^M$, $Q \subseteq Q^M$, $R \subseteq R^M$, and $\mathfrak{I}_\Upsilon(P^M,Q^M,R^M)$ is maximally consistent.

4.4.2 Interpreting the conditional of the ŁK-construction

In this subsection, I will only consider consistent fixed points of Υ.

The conditional modeled by Υ has a philosophically significant algebraic reading: *a tool to compare semantic values* in the sense of Subsection 3.5.1; every consistent fixed point $\mathfrak{I}_\Upsilon(P,Q,R)$ gives conditions for $\varphi \rightarrow \psi$ to have values $1, 0$ *and* $1/2$ in terms of simple relations between the values of φ and ψ in $\mathfrak{I}_\Upsilon(P,Q,R)$.

In building $\mathfrak{I}_\Upsilon(P,Q,R)$, the sets P, Q, R represent a starting distribution of

[28] For a discussion of this point, see [D.A. Martin 2011] and [Burgess 2014].
[29] The proof is routine: see [Cantini 1996], Theorem 30.13 and Lemma 31.2.

values 1, 0, and 1/2, thus giving a partial ordering to $\mathscr{L}_T^{\rightarrow}$-sentences via the corresponding numerical ordering of such values. The first stage of the ŁK-construction gives value 1 to sentences $\varphi \rightarrow \psi$ s.t. $\mathscr{P}_{PQR}(\varphi) \leq \mathscr{P}_{PQR}(\psi)$, or $\mathscr{P}_{PQR}(\varphi) = 0$, or $\mathscr{P}_{PQR}(\psi) = 1$; value 0 to sentences $\varphi \rightarrow \psi$ s.t. $\mathscr{P}_{PQR}(\varphi) > \mathscr{P}_{PQR}(\psi)$ and the value difference is 1; value 1/2 to sentences $\varphi \rightarrow \psi$ s.t. $\mathscr{P}_{PQR}(\varphi) > \mathscr{P}_{PQR}(\psi)$ and the value difference is 1/2. New sets $P^{\mathrm{T}}, Q^{\mathrm{F}}, R^{\mathrm{G}}$ are obtained, that extend the previous ordering of sentences. This process is reiterated: at the second stage, we have also value comparisons between sentences evaluated at the first stage, i.e. value comparisons of value comparisons, and so on. As the stages progress, we evaluate more and more embedded conditionals: each conditional that gets a value is the result of a value comparison, at the appropriate level. At the fixed point, we exhausted the starting information given by P, Q, and R, determining via $\rightarrow$ all the value relations deriving from the initial distribution. *All* these value relations are evaluated at a fixed point because of the fixed-point property: any application of Υ to a fixed-point triple yields that triple, and every conditional that can be evaluated, given the starting distribution, has already a value. Monotonicity ensures that, once a value comparison is established, it does not change, thus the fact that the stages of the construction perform value comparisons transfers to the fixed point.

This table shows the values of conditionals in consistent fixed points of Υ.

$\rightarrow$	0	1/2	1	**N**
0	1	1	1	1
1/2	1/2	1	1	**N**
1	0	1/2	1	**N**
N	**N**	**N**	1	**N**

Table 4.1: $\rightarrow$-comparisons of values in the ŁK-construction (plus **N**)

Let me emphasize that **N** is not a semantic value given by the ŁK-construction, as it indicates a lack of value in a consistent fixed point of Υ. It's status, then, is similar to the status of value $1/2$ when it is associated with the C-gap of a consistent Kripke fixed point. Therefore, in the following, when referring to the values given by the ŁK-construction I will always mean 1, 0, and $1/2$.

Using Table 4.1, we can argue that the ŁK-construction improves on Kripke's theory and on the problems (a)-(e) raised in conjunction with Field's theory in Subsection 3.5.1 (see also Table 3.1).

- Unlike Kripke's theory, the ŁK-construction can employ a non-classical value and a notion of gappiness to evaluate conditionals. A Kripke fixed point only deems equivalent two sentences if they get both value 1 or both value 0, but a fixed point of Υ also deems equivalent its $1/2$-valued sentences.

- As seen in Subsection 3.5.1, many problems of Field's theory are due to the behavior of conditionals getting value different from **1**. In fixed points of Υ, instead, conditionals that get a value different from 1 get a value that measures how close the difference of the values of antecedent and consequent comes to a 1-valued conditional. If $\mathscr{P}_{PQR}(\varphi) = 1$, $\mathscr{P}_{PQR}(\psi_1) = 0$, $\mathscr{P}_{PQR}(\psi_2) = 1/2$, then φ has a greater value than ψ_2 and a still greater value than ψ_1, i.e. $\mathscr{P}_{PQR}(\varphi \to \psi_1) \lneqq \mathscr{P}_{PQR}(\varphi \to \psi_2)$. The model recognizes and expresses this difference:

$$\mathscr{P}_{PQR}(\varphi \to \psi_2) \to (\varphi \to \psi_1)) = 1/2 \tag{4.2}$$

  Contrast this fact with the general failure of comparability in Field's theory, expressed by (3.5). Thus, for every consistent fixed point $\mathfrak{I}_\Upsilon(P,Q,R)$, the Łukasiewicz 3-valued conditional provides a complete map of all the value relations between sentences getting a value in $\mathfrak{I}_\Upsilon(P,Q,R)$. And this is very relevant in the light of the general interest that value relations have for the semantics of $\mathscr{L}_T^{\rightarrow}$, as it was emphasized at the beginning of Subsection 3.5.1.

- Unlike Field's theory, the ŁK-construction provides a simple understanding of semantic values. Not only, 1, 0, and 1/2 retain their usual numerical meaning, but every value plus **N** may result from evaluating a conditional in the ŁK-construction. We can use the value 1/2 of the ŁK-construction to interpret any kind of sentence, including conditionals (unlike Field's theory). As a result, the value 1/2 of the ŁK-construction allows for a more uniform treatment of paradoxical sentences: a simple but striking example is Curry's sentence. The fixed point $\mathfrak{I}_{\Upsilon}(\emptyset,\emptyset,\{\kappa\})$ is consistent and $\kappa \leftrightarrow \neg\kappa$ has value 1 in it. Moreover, also $\mathfrak{I}_{\Upsilon}(\emptyset,\emptyset,\{\kappa,\lambda\})$ is consistent, so both $\lambda \leftrightarrow \neg\lambda$ and $\kappa \leftrightarrow \neg\kappa$ can jointly have value 1, and the ŁK-construction delivers a simple, unified, and numerical treatment of intuitively related paradoxes. Note moreover that in every consistent fixed point of Υ, for every $\varphi \in \mathscr{L}_T^{\rightarrow}$, the sentence $\neg\varphi$ has the same value of the sentence $\varphi \rightarrow 0 = 1$, while Field's theory does not validate this general equivalence.[30]

- The interaction of semantic values with the lack of value (**N**) is also quite intuitive: since **N** is, essentially, the analogue of a Kripkean C-gap for the ŁK-construction, it also interacts with the values in a way similar to C-gaps, according to the strong Kleene paradigm of partiality that I adopted. Such paradigm provides a rationale to identify those value comparisons that can be performed also when some sentence has no semantic value.

- Since when $\varphi \rightarrow \psi$ has no semantic value there is no asking about a value comparison between φ and ψ, it is natural to compare **N** to the sub-space **E** of Field's value space **F**. However, in every consistent fixed point of Υ,

[30] See item (a) in Subsection 3.5.1. There are good reasons to accept this equivalence, such as its classical validity. If one agrees on this equivalence and interprets naïvely the truth predicate, it is natural to see λ and κ as equivalent between them, as λ is equivalent to $\neg\lambda$ and κ is equivalent to $\kappa \rightarrow 0 = 1$ (*modulo* intersubstitutivity of truth).

we can always distinguish between conditionals that are given a value and are treated as in Łukasiewicz 3-valued semantics, and conditionals that are not. The value zone **E**, instead, has a much more complex and less intuitive behavior, as we have seen in Subsection 3.5.1, items (a) and (b).

- The Łukasiewicz 3-valued interpretation, as we have seen, is too strong to be applied to every conditional sentence, and a residual of unevaluated sentences is bound to be generated by every consistent fixed point of Υ. The ŁK-construction, however, shows that, despite the non-monotonic character of the Łukasiewicz conditional, a partial version of this connective can be used within the framework of inductive definitions *a là* Kripke. So, the general picture given by the ŁK-construction is not conceptually different from what we have in Kripke's theory, but it has been shown to provide a *substantial incremental progress* on that construction.

4.4.3 Conditional introduction in the ŁK-construction

The ŁK-construction improves on Kripke's and Field's theories on the introduction of the conditional. Note in the first place that every consistent fixed point $\mathfrak{I}_\Upsilon(P,Q,R)$ has the following closure properties: for all $\varphi, \psi \in \mathscr{L}_T^{\rightarrow}$,

- If $\varphi \in P_\infty$ and $\varphi \rightarrow \psi \in P_\infty$, then $\psi \in P_\infty$.

- If $\varphi \in Q_\infty$, or $\psi \in P_\infty$, or both $\varphi, \psi \in R_\infty$, then $\varphi \rightarrow \psi \in P_\infty$.

The second item can be used to formulate a rule of conditional introduction that is stronger than what is available in Kripke's theory and that expresses genuinely the difference between *conditional assertion* and *assertion of a conditional* in the context of the ŁK-construction.[31] The next definition and theorem, together with the subsequent remarks, establish this point.

[31] See Definition 2.7.1 in Section 2.7.

Definition 4.4.4. (ŁK-logical consequence)

Let $\mathcal{U}$ be any non-empty set of consistent fixed points of Υ. Define the *ŁK-logical consequence generated by* $\mathcal{U}$, in symbols $\models^{\mathcal{U}}_{\text{ŁK}}$, as follows:

$$\Gamma \models^{\mathcal{U}}_{\text{ŁK}} \psi \text{ if and only if } \text{ for every } \mathfrak{I}_\Upsilon(P,Q,R) \in \mathcal{U},$$
$$\text{if for every } \varphi \in \Gamma, \varphi \in P_\infty, \text{ then } \psi \in P_\infty.$$

▲

Theorem 4.4.5.

Let $\mathcal{W}(\varphi,\psi)$ abbreviate the sentence $(\varphi \vee \neg\varphi) \vee \psi \vee [(\varphi \leftrightarrow \neg\varphi) \wedge (\psi \leftrightarrow \neg\psi)]$. For every $\mathcal{U}$ as in Definition 4.4.4, every $\varphi,\psi \in \mathscr{L}_T^{\rightarrow}$, and every $\Gamma \subseteq SENT_{\mathscr{L}_T^{\rightarrow}}$:

$$\Gamma \models^{\mathcal{U}}_{\text{ŁK}} \varphi \rightarrow \psi \quad \textit{if and only if} \quad \begin{cases} \Gamma, \varphi \models^{\mathcal{U}}_{\text{ŁK}} \psi \textit{ and} \\ \quad \Gamma \models^{\mathcal{U}}_{\text{ŁK}} \mathcal{W}(\varphi,\psi). \end{cases}$$

Proof.

I suppress Γ for simplicity. Let's do the left-to-right direction first. Note that, if $\models^{\mathcal{U}}_{\text{ŁK}} \varphi \rightarrow \psi$, then $\varphi \models^{\mathcal{U}}_{\text{ŁK}} \psi$. Let $\mathfrak{I}_\Upsilon(P,Q,R)$ be in $\mathcal{U}$. Then $\varphi \rightarrow \psi \in P_\infty$, i.e. $\varphi \in Q_\infty$, or $\psi \in P_\infty$, or $\varphi,\psi \in R_\infty$. In the first case $\varphi \vee \neg\varphi \in P_\infty$, in the second case $\psi \in P_\infty$, and in the third case $[(\varphi \leftrightarrow \neg\varphi) \wedge (\psi \leftrightarrow \neg\psi)] \in P_\infty$: either way, $\mathcal{W}(\varphi,\psi) \in P_\infty$. Since $\mathfrak{I}_\Upsilon(P,Q,R)$ is arbitrary, this establishes the claim.

As to the other direction, let $\varphi,\psi \in \mathscr{L}_T^{\rightarrow}$ be s.t. $\varphi \models^{\mathcal{U}}_{\text{ŁK}} \psi$ and $\models^{\mathcal{U}}_{\text{ŁK}} \mathcal{W}(\varphi,\psi)$, and consider the following cases.

(A) There is a $\mathfrak{I}_\Upsilon(P',Q',R') \in \mathcal{U}$ s.t. $\psi \in P'_\infty$. In this case, $\varphi \rightarrow \psi \in P'_\infty$.

(B) There is a $\mathfrak{I}_\Upsilon(P^\dagger,Q^\dagger,R^\dagger) \in \mathcal{U}$ s.t. $\varphi \vee \neg\varphi \in P^\dagger_\infty$. In this case, $\varphi \in P^\dagger_\infty$ or $\neg\varphi \in P^\dagger_\infty$. If $\neg\varphi \in P^\dagger_\infty$, then $\varphi \rightarrow \psi \in P^\dagger_\infty$ by the definition of Υ. If $\varphi \in P^\dagger_\infty$, by our assumption that $\varphi \models^{\mathcal{U}}_{\text{ŁK}} \psi$, then also $\psi \in P^\dagger_\infty$, so $\varphi \rightarrow \psi \in P^\dagger_\infty$.

(C) There is a $\mathfrak{I}_{\Upsilon}(P^{\ddagger}, Q^{\ddagger}, R^{\ddagger}) \in \mathcal{U}$ s.t. $[(\varphi \leftrightarrow \neg\varphi) \wedge (\psi \leftrightarrow \neg\psi)] \in P^{\ddagger}_{\infty}$. So, $\varphi, \psi \in R^{\ddagger}_{\infty}$ by Lemma 4.3.6 ($\mathfrak{I}_{\Upsilon}(P^{\ddagger}, Q^{\ddagger}, R^{\ddagger})$ is consistent) and $\varphi \to \psi \in P^{\ddagger}_{\infty}$.

By the assumption that $\models^{\mathcal{U}}_{\text{ŁK}} \mathcal{W}(\varphi, \psi)$, at least one of (A), (B), or (C) is the case, so $\models^{\mathcal{U}}_{\text{ŁK}} \varphi \to \psi$ holds. □

Let **(ŁK→-Intro)** indicate the conditional introduction rule given by the right-to-left direction of Theorem 4.4.5. This rule is interesting under many respects:

1. **(ŁK→-Intro)** is arguably better than all the rules having the form of (2.1) and that are reasonably associated with Kripke's theory, namely "if $\varphi \models^{\mathcal{A}}_{K} \psi$ and $\models^{\mathcal{A}}_{K} \varphi \vee \neg\varphi$, then $\models^{\mathcal{A}}_{K} \varphi \supset \psi$".[32] Unlike this rule, **(ŁK→-Intro)** can use information on sentences having value 1/2. In Kripke's theory, the impossibility to use non-classical values results from the status of C-gaps: the problem is avoided here thanks to P-gaps. As a result, **(ŁK→-Intro)** approximates classical conditional introduction better than the rule just mentioned for Kripke's theory: there are $\mathscr{L}^{\to}_{T}$-sentences φ, ψ s.t. even if $\varphi \models^{\mathcal{A}}_{K} \psi$ and $\varphi \models^{\mathcal{U}}_{\text{ŁK}} \psi$ for every set $\mathcal{A}$ of consistent Kripke fixed points and every set $\mathcal{U}$ of consistent fixed points of Υ, there is no set $\mathcal{A}^*$ of consistent Kripke fixed points s.t. $\models^{\mathcal{A}^*}_{K} \varphi \supset \psi$, whereas there are sets $\mathcal{U}^*$ of consistent fixed points of Υ s.t. $\models^{\mathcal{U}^*}_{\text{ŁK}} \varphi \to \psi$ (e.g. letting φ and ψ be λ).[33]

2. **(ŁK→-Intro)** improves also on rule (3.6), used by Field to introduce conditionals, i.e. "if $\varphi \models_F \psi$ and $\models_F \varphi \vee \neg\varphi$, then $\models_F \varphi \to \psi$". In the first place, as mentioned above, **(ŁK→-Intro)** uses also non-classical values whereas

[32] $\mathcal{A}$ is a non-empty set of consistent Kripke fixed points: see Definition 2.5.7. The introduction of conditionals in Kripke's theory is discussed in Subsection 2.5.2.

[33] The claim that **(ŁK→-Intro)** "approximates classical conditional introduction" better than the rule given for Kripke's theory means that the semantic inferences that can be turned into conditionals in Kripke's theory are a proper subset of the semantic inferences that can be turned into conditionals in the ŁK-construction – for the conditionals connectives available to these theories. This claim does not clash with the fact that **(ŁK→-Intro)** includes the condition $[(\varphi \leftrightarrow \neg\varphi) \wedge (\psi \leftrightarrow \neg\psi)]$, which, were the biconditional $\leftrightarrow$ interpreted classically, would be a contradiction: $\to$, as interpreted by the ŁK-construction, is not the classical conditional.

Field's rule does not. More importantly, however, no characterization of the conditionals $\varphi \to \psi$ such that $\models_F \varphi \to \psi$ holds can be given using Field's rule (3.6). In other words, no analogue of Theorem 4.4.5 holds for Field's rule (3.6), unlike for **(ŁK→-Intro)**.[34] On the other hand, $\models^{\mathcal{U}}_{\text{ŁK}} \mathcal{W}(\varphi, \psi)$ is a *genuine* conditional introduction condition, in the sense of Definition 2.7.1, since for every set $\mathcal{U}$ of consistent fixed points of Υ, the schema $\mathcal{W}(\varphi, \psi)$ is not equivalent to the $\models^{\mathcal{U}}_{\text{ŁK}}$-validity of all the corresponding conditionals (e.g. $\models^{\mathcal{U}}_{\text{ŁK}} \mathcal{W}(0 = 0, \lambda)$, but $\not\models^{\mathcal{U}}_{\text{ŁK}} 0 = 0 \to \lambda$), and the same holds of all its subschemata, as the proof of Theorem 4.4.5 shows. Thus, in the ŁK-construction we have a schema, namely $\mathcal{W}(\varphi, \psi)$, that expresses genuinely the difference between conditional assertion and assertion of a conditional for the semantics given by fixed points of Υ, overcoming one of the major difficulties of Field's theory.[35] Moreover, **(ŁK→-Intro)** indicates an interesting conceptual link between conditional assertion and assertion of a conditional: it is consistent with weak naïveté to validate a semantic inference in the object-language, via a conditional, not only when the sentences involved have classical values (value 1 for the consequent), but also when they have a positive gappy status, i.e. value $1/2$ of Łukasiewicz 3-valued semantics rendered partial in the strong Kleene paradigm.

Finally, it is immediate to see that *modus ponens* holds for every notion of logical consequence $\models^{\mathcal{U}}_{\text{ŁK}}$ given as in Definition 4.4.4. Moreover, the introduction and elimination rules for $\neg$, $\wedge$, $\vee$, and $\forall$ in $\models^{\mathcal{U}}_{\text{ŁK}}$ are the same of Kripke's theory, and thus approximate closely their classical counterparts.

Remark 4.4.6.

Definition 4.4.4 parallels Definition 2.5.7 and it has the same cost of excluding

[34] The introduction of conditionals in Field's theory was discussed in Subsection 3.5.2.

[35] See the discussion in Subsection 3.5.2.

inconsistent fixed points. This restriction is motivated as in Kripke's theory and is discussed in Remark 2.5.8, to which I refer the reader, since the points made there apply, *mutatis mutandis*, here. In short, had I accepted arbitrary fixed points of Υ in Definition 4.4.4, several basic pieces of semantic reasoning would have been lost: e.g. there are sets $\mathcal{U}'$ of fixed points of Υ and sentences $\varphi, \psi \in \mathscr{L}_T^{\rightarrow}$ s.t. $\models^{\mathcal{U}'}_{\text{ŁK}} \varphi \to \psi$ but $\varphi \not\models^{\mathcal{U}'}_{\text{ŁK}} \psi$. ▲

4.4.4 Summing up: general aspects of the ŁK-construction

Similarly to Kripke's theory, the ŁK-construction is a general template to give a fairly rich array of interpretations of $\mathscr{L}_T^{\rightarrow}$, the fixed points of Υ, that are good candidates to improve on the weaknesses of Kripke's theory without the problems of Field's theory. In fact, every fixed point of Υ:

(i) preserves Kripke's theory, in the strong sense given by Corollary 4.4.2;

(ii) gives a simple and philosophically significant reading to the conditional, which can be seen as a tool to compare semantic values 1, 0, *and the non-classical value* $1/2$ (in the sense of Subsections 3.5.1 and 4.4.2), being a partial version of Łukasiewicz 3-valued conditional;

(iii) is given using the relatively simple means of inductive definitions, the same means employed by Kripke's construction, being therefore acceptable for anyone accepting Kripke's theory.

In addition, the fixed points of Υ that have a non-empty starting hypothesis for P-gappy sentences improve essentially on Kripke's theory, as they:

(i+) add to Kripke's theory a valuation for P-gappy sentences that makes explicit the notion of gappiness we want to capture – and this improves on Kripke's treatment of truth-theoretic facts (as shown in Section 4.3);

(ii+) allow us to formalize and validate claims that Kripke's theory cannot give (e.g. statement (G)), yield introduction/elimination rules that use also non-classical information (and express genuinely the difference between conditional assertion and assertion of the conditional), and define a determinateness operator that is stronger than strong Kleene negation.

These general features show that the ŁK-construction is a promising framework to address FIELD'S PROGRAM. This construction preserves Kripke's theory (Subsection 4.4.1) and endows it with a conditional that is well-behaved according to criteria **(W1)**, **(W2)**, and **(W3)**.[36] As for **(W1)**, the conditional interpreted by the ŁK-construction has a simple algebraic reading (Subsection 4.4.2), that turns out to be philosophically quite significant. As for **(W2)**, the ŁK-construction features a genuine conditional introduction that approximates the classical rule of conditional introduction in a nice way (Subsection 4.4.3). As for **(W3)**, the situation is less clear, however one could say that the ŁK-construction can be used to model several notions of gappiness (Subsection 4.4.1), and its conditional behaves uniformly in every application (e.g. defining a notion of determinateness, and so on), going some way towards accommodating **(W3)** as well, at least to an extent.

It is clear, however, that concrete progress in addressing FIELD'S PROGRAM can only be given by specific fixed points of Υ. I turn now to an application of the ŁK-construction that arguably provides a nice option to deal with FIELD'S PROGRAM, a philosophically motivated collection of fixed points of Υ.

4.5 Applications and extensions

I will now isolate some fixed points of Υ that embody interesting views on gappiness and on the role of the conditional. First, I will present a family of fixed

[36] They were recalled in footnote 1, Section 4.1.

points of Υ that characterize a *meta-theoretic view* of the role of the conditional in $\mathscr{L}_T^{\rightarrow}$, constituting what is arguably a completion of Kripke's theory. Second, I will consider a variant of this application, that could open the way to axiomatic theories designed along the lines of the ŁK-construction.

4.5.1 $\mathscr{L}_T$-grounded conditionals

Consider the distributions of values obtained (*not explicitly*) in a consistent Kripke fixed points *for the language* $\mathscr{L}_T$ and the following kind of fixed points.

Definition 4.5.1. (CAB fixed points)
For $A, B \subseteq SENT_{\mathscr{L}_T}$ s.t. $\mathfrak{I}_\Phi(A,B)$ is consistent, put $C^{\mathrm{A,B}} := SENT_{\mathscr{L}_T} \setminus (\mathfrak{E}_\Phi(A,B) \cup \mathfrak{A}_\Phi(A,B))$. Call a fixed point of the form $\mathfrak{I}_\Upsilon(A,B,C^{\mathrm{A,B}})$ a CAB *fixed point*. ▲

In other words, CAB fixed points exclude from their starting hypotheses (codes of) sentences that belong to $\mathscr{L}_T^{\rightarrow}$ but not to $\mathscr{L}_T$. This entails that some sentences are not in any CAB fixed point because of the connectives or terms they feature, such as κ or $T\ulcorner\kappa\urcorner$, but this is in line with the rationale of these fixed points: they ignore sentences that are not *grounded*, in the Kripkean sense, *in the conditional-free language* $\mathscr{L}_T$.[37] $C^{\mathrm{A,B}}$ is an important kind of set in a partial semantics. Kripke's theory was conceived for $\mathscr{L}_T$, the sublanguage of $\mathscr{L}_T^{\rightarrow}$ that does not have $\rightarrow$. Assuming $C^{\mathrm{A,B}}$, we make available to the broader construction via Υ *all* the negative information relative to C-Gaps in $\mathscr{L}_T$, integrating it in P-Gaps.

For simplicity, in this section I will always use A and B as in Definition 4.5.1, and I will write C for $C^{\mathrm{A,B}}$: this should not cause any confusion, since the sets A and B used to define $C^{\mathrm{A,B}}$ will always be clear from the context.

CAB fixed points have nice features: to begin with, they are always consistent.

[37] I am referring primarily to Kripke's informal idea of groundedness (see [Kripke 1975], p. 694 and p. 701), but the present discussion could also be related to the more formal notion that Kripke discusses on p. 706 and following. For an analysis of Kripkean groundedness, see [Yablo 1982].

Theorem 4.5.2.

Every CAB *fixed point is consistent.*

Proof.

Let $\mathfrak{I}_{\Upsilon}(A,B,C)$ be a CAB fixed point. The proof, by induction on the build-up of $\mathfrak{I}_{\Upsilon}(A,B,C)$, is made of distinct but similar claims, and I will do only some of them. Let "pwd" abbreviate "pairwise disjoint".

1 A^0, B^0 and C^0 are pwd. Immediate by definition of C.

2a For every ordinal α, if A^α, B^α, C^α are pwd, $\mathfrak{E}_\Phi(A^\alpha,B^\alpha)\cap\mathfrak{A}_\Phi(A^\alpha,B^\alpha)=\emptyset$. If $\alpha=0$, it's immediate by claim **1**. Let $\alpha>0$ and suppose by contradiction that for a $\varphi\in\mathscr{L}_T^{\rightarrow}$, $\varphi\in\mathfrak{E}_\Phi(A^\alpha,B^\alpha)\cap\mathfrak{A}_\Phi(A^\alpha,B^\alpha)$. So, there is a least $\delta\geq 0$ s.t.: $\varphi\notin\mathfrak{E}^\delta_\Phi(A^\alpha,B^\alpha)\cap\mathfrak{A}^\delta_\Phi(A^\alpha,B^\alpha)$ but $\varphi\in\mathfrak{E}^{\delta+1}_\Phi(A^\alpha,B^\alpha)\cap\mathfrak{A}^{\delta+1}_\Phi(A^\alpha,B^\alpha)$. Let's argue by cases according to the logical form of φ. Induction hypothesis: for all $\gamma\leq\delta$, $\mathfrak{E}^\gamma_\Phi(A^\alpha,B^\alpha)$, $\mathfrak{A}^\gamma_\Phi(A^\alpha,B^\alpha)$, and $\mathfrak{H}^\gamma_\Psi(C^\alpha,A^\alpha,B^\alpha)$ are pwd.

φ is $(s=t)$ for $s,t\in TER_{\mathscr{L}_T^{\rightarrow}}$. Then $\varphi\in\mathfrak{E}^{\delta+1}_\Phi(A^\alpha,B^\alpha)\cap\mathfrak{A}^{\delta+1}_\Phi(A^\alpha,B^\alpha)$ iff s and t both coincide and differ in their value, which is absurd.

φ is $\neg\psi$. So $\psi\in\mathfrak{E}^\delta_\Phi(A^\alpha,B^\alpha)\cap\mathfrak{A}^\delta_\Phi(A^\alpha,B^\alpha)$, against the IH.

φ is $\psi\wedge\chi$. So $\psi\in\mathfrak{E}^\delta_\Phi(A^\alpha,B^\alpha)$, $\chi\in\mathfrak{E}^\delta_\Phi(A^\alpha,B^\alpha)$ but also $\psi\in\mathfrak{A}^\delta_\Phi(A^\alpha,B^\alpha)$ or $\chi\in\mathfrak{A}^\delta_\Phi(A^\alpha,B^\alpha)$, against the IH.

φ is $\psi\vee\chi$. Dual to the above case.

φ is $\psi\rightarrow\chi$. There is no clause for $\rightarrow$ in Φ, so $\varphi\in\mathfrak{E}^{\delta+1}_\Phi(A^\alpha,B^\alpha)\cap\mathfrak{A}^{\delta+1}_\Phi(A^\alpha,B^\alpha)$ iff $\varphi\in\mathfrak{E}^\delta_\Phi(A^\alpha,B^\alpha)\cap\mathfrak{A}^\delta_\Phi(A^\alpha,B^\alpha)$, against the IH.

φ is $\forall x\chi(x)$. So, for all $t\in CTER_{\mathscr{L}_T^{\rightarrow}}$, $\chi(t)\in\mathfrak{E}^\delta_\Phi(A^\alpha,B^\alpha)$ but there is a $s\in CTER_{\mathscr{L}_T^{\rightarrow}}$ such that $\chi(s)\in\mathfrak{A}^\delta_\Phi(A^\alpha,B^\alpha)$, against the IH.

φ is Tt. So, $t=\ulcorner\chi\urcorner$, $\chi\in\mathfrak{E}^\delta_\Phi(A^\alpha,B^\alpha)\cap\mathfrak{A}^\delta_\Phi(A^\alpha,B^\alpha)$, against the IH.

2b For every ordinal α, if A^α, B^α, C^α are pwd, $\mathfrak{E}_\Phi(A^\alpha, B^\alpha) \cap \mathfrak{H}_\Psi(C^\alpha, A^\alpha, B^\alpha) = \emptyset$. The proof is similar to the proof of claim **2a**.

2c For every ordinal α, if A^α, B^α, C^α are pwd, $\mathfrak{A}_\Phi(A^\alpha, B^\alpha) \cap \mathfrak{H}_\Psi(C^\alpha, A^\alpha, B^\alpha) = \emptyset$. The proof is entirely dual to the proof of claim **2a**.

3a For every ordinal α, if A^α, B^α, and C^α are pwd, $A^{\alpha+1} \cap B^{\alpha+1} = \emptyset$. See the proof of claim **3b** below.

3b For every ordinal α, if A^α, B^α, and C^α are pwd, then $A^{\alpha+1} \cap C^{\alpha+1} = \emptyset$.

Suppose that $A^{\alpha+1} \cap C^{\alpha+1} \neq \emptyset$. $\mathfrak{E}_\Phi(A^\alpha, B^\alpha)$, $\mathfrak{A}_\Phi(A^\alpha, B^\alpha)$, $\mathfrak{H}_\Psi(C^\alpha, A^\alpha, B^\alpha)$ are pwd by claims **2a**, **2b**, and **2c**, so there is a $\psi \to \chi$ s.t. $\psi \to \chi \in A^{\alpha+1} \cap C^{\alpha+1} \setminus A^\alpha \cap C^\alpha$. So, one of the following obtains:

$$\text{(I)}\ \ \psi \in \mathfrak{A}_\Phi(A^\alpha, B^\alpha) \text{ and } \begin{cases} \psi \in \mathfrak{E}_\Phi(A^\alpha, B^\alpha) \text{ and } \chi \in \mathfrak{H}_\Psi(C^\alpha, A^\alpha, B^\alpha), \text{ or} \\ \psi \in \mathfrak{H}_\Psi(C^\alpha, A^\alpha, B^\alpha) \text{ and } \chi \in \mathfrak{A}_\Phi(A^\alpha, B^\alpha) \end{cases}$$

$$\text{(II)}\ \ \chi \in \mathfrak{E}_\Phi(A^\alpha, B^\alpha) \text{ and } \begin{cases} \psi \in \mathfrak{E}_\Phi(A^\alpha, B^\alpha) \text{ and } \chi \in \mathfrak{H}_\Psi(C^\alpha, A^\alpha, B^\alpha), \text{ or} \\ \psi \in \mathfrak{H}_\Psi(C^\alpha, A^\alpha, B^\alpha) \text{ and } \chi \in \mathfrak{A}_\Phi(A^\alpha, B^\alpha) \end{cases}$$

$$\text{(III)}\ \ \psi, \chi \in \mathfrak{H}_\Psi(C^\alpha, A^\alpha, B^\alpha) \text{ and } \begin{cases} \psi \in \mathfrak{E}_\Phi(A^\alpha, B^\alpha), \chi \in \mathfrak{H}_\Psi(C^\alpha, A^\alpha, B^\alpha), \text{ or} \\ \psi \in \mathfrak{H}_\Psi(C^\alpha, A^\alpha, B^\alpha), \chi \in \mathfrak{A}_\Phi(A^\alpha, B^\alpha) \end{cases}$$

But any of (I), (II), and (III) contradicts our assumption and **2a**, **2b**, and **2c**.

3c For every ordinal α, if A^α, B^α, and C^α are pwd, then $B^{\alpha+1} \cap C^{\alpha+1} = \emptyset$. The proof is entirely dual to the proof of claim **3b**.

4 For every limit ordinal γ, A^γ, B^γ, and C^γ are pwd.

This is immediate from **3a**, **3b**, and **3c**: Υ is monotone, so for γ limit:

$$\langle A^\gamma, B^\gamma, C^\gamma \rangle = \Upsilon^\gamma(A, B, C) = \bigcup_{\delta < \gamma} \Upsilon^\delta(A, B, C) = \bigcup_{\delta < \gamma} \langle A^\delta, B^\delta, C^\delta \rangle,$$

where A^δ, B^δ, and C^δ are pwd for all $\delta < \gamma$ by IH.

□

CAB fixed points confer a real expressive capability to $\rightarrow$, counterbalancing the triviality of the fixed points presented in Corollary 4.4.2.

Lemma 4.5.3. (Non-triviality)

For every CAB *fixed point* $\mathfrak{I}_\Upsilon(A,B,C)$ *and every* $n \in \omega$*, there exist* $\mathscr{L}_T^{\rightarrow}$*-sentences* φ, ψ *s.t.* $\varphi \rightarrow \psi \in A^{n+1} \setminus A^n$ *but* $\neg\varphi \vee \psi \in C^{n+1}$.

Proof.

Let $\mathfrak{I}_\Upsilon(A,B,C)$ be a CAB fixed point, and let χ be a sentence in C. At least one such χ exists, since $\mathfrak{I}_\Phi(A,B)$ is consistent, and so $C \neq \emptyset$. Define the following sentences by primitive recursion:

$$\begin{cases} \varphi_0 \text{ is } \chi \\ \varphi_{n+1} \text{ is } 0=0 \rightarrow \varphi_n \end{cases}$$

By induction, for every $n \in \omega$, $\varphi_n \rightarrow \varphi_n \in A^{n+1} \setminus A^n$ but $\neg\varphi_n \vee \varphi_n \in C^{n+1}$. □

CAB fixed points give a *non-trivial* treatment to the conditional, in the sense that there are conditionals that are given value 1 or 0 at high stages of the construction of the fixed point, whereas their $\supset$-version keeps value $1/2$.

Remark 4.5.4.

Non-triviality holds at infinite stages too. Here is a simple example. Let $\mathfrak{I}_\Upsilon(A,B,C)$ and χ be as above, and define the following primitive recursive function t:

$$\begin{cases} t(0, \ulcorner\chi\urcorner) := \chi \\ t(n+1, \ulcorner\chi\urcorner) := T\ulcorner 0=0 \rightarrow \underset{\cdot}{t}(n, \ulcorner\chi\urcorner)\urcorner \end{cases}$$

It's straightforward to verify that:

- $(\forall n)[T\ulcorner \underset{\cdot}{t}(n,\ulcorner\chi\urcorner)\urcorner] \to (\forall n)[T\ulcorner \underset{\cdot}{t}(n,\ulcorner\chi\urcorner)\urcorner] \in A^{\omega+2} \setminus A^{\omega+1}$
- $\neg(\forall n)[T\ulcorner \underset{\cdot}{t}(n,\ulcorner\chi\urcorner)\urcorner] \vee (\forall n)[T\ulcorner \underset{\cdot}{t}(n,\ulcorner\chi\urcorner)\urcorner] \in C^{\omega+2} \setminus C^{\omega+1}$

In fact, new conditionals enter in CAB fixed points at ordinal stages much larger than $\omega+2$ (but smaller than ω_1^{CK}).[38] ▲

As for the expression of truth-theoretic facts, Theorem 4.3.1, Corollaries 4.3.2, 4.3.4, 4.3.5, and Lemma 4.3.6 hold for every CAB fixed point as well. The divergence between weak and strong naïveté (Theorem 4.3.1 and Corollary 4.3.4) is now interpreted in the light of the role that CAB fixed points confer to the conditional. This application of the ŁK-construction embodies a sort of *meta-theoretic view* of the conditional, which is explicitly brought to the object-language level by CAB fixed points – of course, the conditional retains also its simple reading of a tool to compare semantic values given in general by the ŁK-construction. CAB fixed points are designed to formalize and validate *exactly* the conditionals and equivalences between $\mathscr{L}_T$-sentences that Kripke's construction cannot in general validate, plus the derived conditionals and equivalences between $\mathscr{L}_T^{\to}$-sentences obtained starting from an original Kripkean partition.

In this perspective, the language one is really interested in is $\mathscr{L}_T$: the detour through $\mathscr{L}_T^{\to}$ is instrumental to express better the semantics of $\mathscr{L}_T$. And in fact CAB fixed points do a nice job in improving on the problems that Kripke's theory shows in accounting for the semantics of $\mathscr{L}_T^{\to}$ and that were summarized at the beginning of Section 2.6, labelled as **(K1)**, **(K2)**, **(K3)**.[39] As to **(K1)**, admittedly,

[38] I do not formulate a general version of Lemma 4.5.3, as it would require to deal with some complicated and ultimately not relevant issues on ordinal notations and it would lead me away from the conceptual points I want to make. It seems sufficient to have shown that new conditionals are added as the stages of the build-up of CAB fixed points progress, including infinite stages.

[39] Let me recall them here for the reader's convenience.

(K1) In consistent Kripke fixed points no logical laws are validated, and there is no prospect lawlessness is not remedied, but this is understandable in view of the fact that CAB fixed points do not target the entire language $\mathscr{L}_T^{\rightarrow}$ but only the "$\mathscr{L}_T$-grounded" fragment of it; as for conditional introduction, Theorem 4.4.5 applies to every collection of CAB fixed points, thus associating with them a strong and interesting rule for conditional introduction.[40] As to **(K2)**, the Tarski biconditional $\varphi \leftrightarrow T\ulcorner\varphi\urcorner$ holds in a CAB fixed point $\mathfrak{I}_\Upsilon(A,B,C)$ whenever φ is a $\mathscr{L}_T$-sentence of the original Kripkean partition given by A, B, and C, or φ is a $\mathscr{L}_T^{\rightarrow}$-sentence derived from it in the build-up of $\mathfrak{I}_\Upsilon(A,B,C)$. So, the equivalence between φ and $T\ulcorner\varphi\urcorner$ is validated by the theory for every sentence targeted by Kripke's theory, plus all the results of value comparisons between them. As to **(K3)**, the dangerous, liar-like paradoxes that arise in $\mathscr{L}_T$ are also accommodated nicely by CAB fixed points: not only it is possible to classify them explicitly as P-gappy in the fixed-point triples of Υ, we can also classify explicitly all the sentences in the resulting set C_∞ as not determinately true. Finally, the treatment of the liar paradox does justice to the intuition that it is equivalent to its own negation, since every CAB fixed point validates the sentence $\lambda \leftrightarrow \neg\lambda$, which formalizes the statement (G) mentioned in Section 2.5.1. Regrettably, however, truth-teller-like paradoxes receive the same, unsatisfactory treatment they get in Kripke's theory or in Field's theory.

For all these reasons, one could claim that CAB fixed points, almost literally, embody the addition of a new conditional to Kripke's theory for its intended language: they preserve the original Kripke's construction for $\mathscr{L}_T$ and add to it exactly

of recovering an interesting and strong semantic rule for conditional introduction.

(K2) For every sentence φ and every consistent Kripke fixed point $\mathfrak{I}_\Phi(P,Q)$, φ has value 1 (0, 1/2) in $\mathfrak{I}_\Phi(P,Q)$ if and only if $T\ulcorner\varphi\urcorner$ has value 1 (0, 1/2) in it, but this fact cannot be represented within the theory in many cases, including apparently simple ones – inconsistent fixed points are an exception, but they are not interesting for my purposes.

(K3) Kripke's theory fails to deliver a satisfactory account of paradoxes. On the one hand, Kripke's official theory does not include any treatment of the liar, the truth-teller, and other paradoxes. Moreover, some central features of these paradoxes cannot be addressed even enlarging Kripke's theory to include C-gaps.

[40] See Subsection 4.4.3

what was missing to express the semantics of $\mathscr{L}_T$ in a satisfactory way.

An important CAB fixed point is $\mathfrak{I}_\Upsilon(\emptyset,\emptyset,C_{\emptyset,\emptyset})$, i.e. the fixed point of Υ applied to the least Kripke fixed point and its C-Gappy $\mathscr{L}_T$-sentences. For $A,B \subseteq SENT_{\mathscr{L}_T}$, if $\mathfrak{I}_\Upsilon(A,B)$ is consistent, $C_{\mathrm{A,B}} \subseteq C_{\emptyset,\emptyset}$, so this CAB fixed point exploits in full a very rich assumption about sentences given value $1/2$.

Remark 4.5.5.

[Feferman 1984] observes that the consistency of Tarski biconditionals $\varphi \leftrightarrow T\ulcorner\varphi\urcorner$ formulated with Łukasiewicz 3-valued conditional *restricted to* $\varphi \in \mathscr{L}_T$ can be proven along the lines of the consistency proof of similarly restricted instances of naïve comprehension given by [Brady 1971]. The proof of Theorem 4.5.2 gives a very simple demonstration of this fact.

Feferman insists also on the unsatisfactory character of Brady's restriction, urging for some improvement on it. He proposes to supplement classical logic with an *intensional* conditional $\rightarrow^\star$ and its standardly defined biconditional $\leftrightarrow^\star$, which is given a semantics that validates every $\leftrightarrow^\star$-Tarski biconditional.[41] The semantics of $\rightarrow^\star$, however, is quite disappointing: for example, not even *modus ponens* is validated for it.[42] CAB fixed points may be seen as an answer to Feferman's instigation to recover more instances of Tarski biconditionals, with a conceptually motivated restriction of the excluded instances (also less draconian than Brady's one), and a better-behaved conditional. ▲

4.5.2 Provably P-gappy sentences

A variant of the above application is obtained considering the next definition.

Definition 4.5.6.

Consider the axiomatic theory of truth PKF, formulated in $\mathscr{L}_T$, given in [Halbach

[41] This theory is obtained in [Aczel and Feferman 1980].

[42] See [Feferman 1984], §11.

and Horsten 2006], and define the set $C_{\mathrm{PKF}} := \{\varphi \in \mathscr{L}_T \,|\, \mathsf{PKF} \vdash \varphi \Leftrightarrow \neg\varphi\}$.[43] ▲

Lemma 4.5.7.

If $\varphi \in C_{\mathrm{PKF}}$*, then:* (i) $\mathsf{PKF} \nvdash \varphi$ *and* $\mathsf{PKF} \nvdash \neg\varphi$*;* (ii) $\varphi \notin \mathfrak{E}_\Phi(P,Q) \cup \mathfrak{A}_\Phi(P,Q)$ *for all* P*,* $Q \subseteq SENT_{\mathscr{L}_T}$ *s.t.* $\mathfrak{I}_\Phi(P,Q)$ *is consistent.*

Proof.

(i) is immediate: PKF is consistent and closed under (Cut). For (ii), if $\varphi \in C_{\mathrm{PKF}} \cap \mathfrak{E}_\Phi(P^*,Q^*)$ for a consistent fixed point $\mathfrak{I}_\Phi(P^*,Q^*)$ (with $P^*, Q^* \subseteq SENT_{\mathscr{L}_T}$), then $\neg\varphi \in \mathfrak{A}_\Phi(P^*,Q^*)$. If $\varphi \in C_{\mathrm{PKF}}$, then $\mathsf{PKF} \vdash \varphi \Rightarrow \neg\varphi$. But then, by the soundness of PKF, we have that $\neg\varphi \in \mathfrak{E}_\Phi(P^*,Q^*)$, against the consistency of $\mathfrak{I}_\Phi(P^*,Q^*)$.[44] The case of Kripke falsity-sets is similar. □

One could see C_{PKF} as giving a notion of *provable gappiness*, a specific realization of the more general idea of positive gappiness. Lemma 4.5.7 tells us something about what kind of gappiness is identified by PKF: for example, λ is provably gappy in it, τ is not. More importantly, the lemma gives us the next result.

Theorem 4.5.8.

If A*,* $B \subseteq SENT_{\mathscr{L}_T}$ *are s.t.* $\mathfrak{I}_\Phi(A,B)$ *is consistent,* $\mathfrak{I}_\Upsilon(A,B,C_{\mathrm{PKF}})$ *is consistent.*

Proof sketch.

By Lemma 4.5.7, if A, $B \subseteq SENT_{\mathscr{L}_T}$ are s.t. $\mathfrak{I}_\Phi(A,B)$ is consistent, $C_{\mathrm{PKF}} \subseteq C^{\mathrm{A,B}}$. This gives us the base case of an inductive proof that, for the other stages, is substantially identical to the proof of Theorem 4.5.2. □

More generally, all the results holding for CAB fixed points hold for fixed points of the form $\mathfrak{I}_\Upsilon(A,B,C_{\mathrm{PKF}})$, for $A, B \subseteq SENT_{\mathscr{L}_T}$ such that $\mathfrak{I}_\Phi(A,B)$ is consistent. For example, this include non-triviality in the sense of Lemma 4.5.3 and Remark 4.5.4.

[43] Recall that PKF is formulated in a sequent calculus.

[44] For the soundness of PKF, see [Halbach and Horsten 2006], especially the proof of Lemma 20, p. 695; see also the equivalences on the bottom of p. 680.

However, the main point of interest of the restriction to C_{PKF} is that it is a recursively enumerable set, so quite simple. As a consequence, if A and B are inductive, then also $\mathfrak{I}_{\Upsilon}(A,B,C_{\mathrm{PKF}})$ is purely inductive – this is the case, for example, if $A = \emptyset = B$. This fact seems both technically and conceptually important: it shows that it is possible to have an expressive paracomplete theory of naïve truth that features a well-behaved conditional connective *and* that proves to be very simple (for the standards of semantic theories of truth). At the same time, the use of sets such as C_{PKF} prompts us to develop *calculi* to determine positive assertions of gappiness, that could be interpreted by models such as the ŁK-construction. And this, in turn, would be interesting in order to associate a logic with some collection of fixed points of Υ. Corollary 4.4.2 indicates that we do not want to find a calculus that is sound with respect to *all* the fixed points of Υ (this would, essentially, reduce to PKF formulated in $\mathscr{L}_T^{\rightarrow}$). However, we could find a calculus which is sound with respect to some family of fixed points of Υ whose starting sets for P-gappy sentences are non-empty, recursively enumerable sets, just as C_{PKF}.[45]

4.6 Summary of the chapter

In this chapter I presented a construction, called the ŁK-construction, devised to address FIELD'S PROGRAM in a way that is conceptually and technically simple and that avoids the difficulties of Field's theory that were analyzed in Chapter 3.

In Section 4.1, I motivated the approach to FIELD'S PROGRAM I advocate in this chapter, introducing the idea of using Łukasiewicz 3-valued semantics within a Kripke-style construction. In Section 4.2, I described the abstract framework of the

[45]In the case of fixed points extending $\mathfrak{I}_{\Upsilon}(\emptyset,\emptyset,C_{\mathrm{PKF}})$, for example, it is clear how to describe such calculus. The arithmetical and truth-theoretic parts of the theory would be given by the axioms and rules of PKF and its logic would consist of all the axioms and rules of PKF for $\neg$, $\wedge$, $\vee$, and $\forall$. As to the conditional, *modus ponens* would hold, together with the following rule for introducing the conditional (matching **(ŁK$\rightarrow$-Intro)**): the theory proves $\Rightarrow \varphi \rightarrow \psi$ whenever the theory proves $\varphi \Rightarrow \psi$ and one of the following is the case: (i) the theory proves $\Rightarrow \varphi \vee \neg\varphi$, (ii) the theory proves $\Rightarrow \psi$, (iii) the theory proves both $\varphi \Leftrightarrow \neg\varphi$ and $\psi \Leftrightarrow \neg\psi$.

ŁK-construction (Subsection 4.2.1), I gave the details of the construction (Subsection 4.2.2), and I discussed the idea of a paradigm of partiality (Subsection 4.2.3). The properties of the ŁK-construction were reviewed in Section 4.3, while Section 4.4 explored how the ŁK-construction addresses FIELD'S PROGRAM, focusing on the recovery of Kripke's theory (Subsection 4.4.1), on the interpretation of the conditional (Subsection 4.4.2), and on the question of conditional introduction in the ŁK-construction (Subsection 4.4.3), summarizing then the results obtained (Subsection 4.4.4). Finally, Section 4.5 presented two philosophically motivated applications of the ŁK-construction (Subsections 4.5.1 and 4.5.2).

Notice that, so far, I have worked with Łukasiewicz' 3-valued semantics, rendered partial along the lines of the strong Kleene paradigm. It is natural to consider the idea of applying this partiality paradigm to other semantics, that may enable us to evaluate more $\mathscr{L}_T^{\rightarrow}$-sentences. Finitely valued Łukasiewicz semantics with more than 3 values are obvious candidates for this extension. An analogue of the ŁK-construction for even a relatively small number of values, however, would be very cumbersome to set up and to work with, making the resulting theory obscure.[46] Furthermore, the complications would be multiplied if one wanted to adopt, in the framework of the ŁK-construction, the continuum valued Łukasiewicz evaluation scheme within the strong Kleene partiality paradigm, a combination that seems a very interesting option in order to address FIELD'S PROGRAM.

In addition, note that the problems of the treatment of truth-teller-like paradoxes that affect Kripke's and Field's theories affect also the ŁK-construction.

In the next chapter, I will present a semantics that overcomes these difficulties: it allows us to apply the strong Kleene partiality paradigm to *every* Łukasiewicz semantics, and several other evaluation schemes. The resulting theory offers also a more articulate account of paradoxes, including truth-teller-like cases and others.

[46] Writing an analogue of Definition 4.2.9 for Łukasiewicz 7-valued semantics seems already quite intricate, so a much larger number of values would be almost impossible to treat.

Chapter 5

A New Graph-Theoretic Semantics

5.1 Introduction

In this chapter I will provide a semantics for the language $\mathscr{L}_T^{\rightarrow}$ that, I will argue, constitutes a nice proposal to address FIELD'S PROGRAM and to account for truth-theoretic paradoxes. Unlike Kripke's theory, Field's theory, and the ŁK-construction, the theory developed here belongs to a kind of model-theoretic constructions that could be called the group of "graph-theoretic semantics", namely those semantics that use graph-theoretic tools to construct evaluation functions for their target language and to prove their properties. The idea of using graphs to give a semantics for a language featuring a naïve truth predicate and to analyze truth-theoretic paradoxes has been used several times in the truth-theoretic literature.[1] Relating my construction to all the other works in this area I am aware of, however, would require a chapter on its own and it would lead me far away from the

[1]For other graph-theoretic semantics and analyses of paradoxes, see [Davis 1979] (see also [Hazen 1981]), [Yablo 1982], [Barwise and Etchemendy 1987], [Minari 1987], [Gaifman 1988], [Gaifman 1992], [Gaifman 2000] (see also [Whittle 2015]), [Maudlin 2004], [Rabern, Rabern, and Macauley 2013], [Cook 2014], [Hansen 2015].

concerns of this thesis. So, I will occasionally comment on the relation between my theory and other approaches, but I will not attempt a systematic comparison. I hope that my presentation will highlight the specificities of my approach, making it easy to compare it with other semantics for an interested reader.

5.1.1 Semantic graphs: some heuristics

Let me introduce my approach informally. Take a sentence φ of $\mathscr{L}_T^{\rightarrow}$: what information is needed to give it a value? Intuitively, the logical form of φ should indicate this. For example, if φ is an atomic arithmetical sentence, i.e. of the form $s = t$, it is sufficient to check whether s and t have the same value in the standard model $\mathbb{N}$. Assuming a broadly compositional attitude towards evaluation, moreover, it is clear how we should evaluate complex sentences. If φ is a negated sentence $\neg\psi$, in order to assign a value to φ we need to have enough information to assign a value to ψ. The same goes for the other connectives and quantifiers: to give a value to $\psi \wedge \chi$, for example, we must ask for the values of ψ and χ. Actually, adopting the usual reading of $\wedge$ and the strong Kleene paradigm of partiality,[2] if one of ψ and χ has value 0, the value of the other sentence is not needed: still, it is clear that ψ and χ are the sentences to look at in order to assign a value to $\psi \wedge \chi$.

So far, so good. The truth predicate, however, complicates the picture. Since T is given a naïve interpretation, in order to evaluate a sentence of the form $T\ulcorner\psi\urcorner$, we must give a value to ψ first. In some paradoxical cases, however, while searching for enough information to assign a value to ψ, we are required the value of $T\ulcorner\psi\urcorner$ again: when this happens, we ended up in a *loop*. This description of a loop, let me point out, suggests the idea of a *recognizable* circularity: after a *finite* number of steps since the first occurrence of $T\ulcorner\psi\urcorner$, following the unravelling process I just sketched, we would find a second occurrence of $T\ulcorner\psi\urcorner$, deriving from the former.

[2]In the sense of Subsection 4.2.3.

Loops, however, can be seen as a resource rather than as a problem.[3] In fact, loops of the kinds I just described are not useless, as they display some *relational information* about the sentences we find in them. Now, adopting a broadly compositional semantics, several readings of connectives, quantifiers, and T can be expressed by equations in a value space V, expressing relations between the values of two or more sentences. The two following evaluation clauses are a nice example:

$$\text{the value of } \neg\psi = 1 - \text{ the value of } \psi,$$
$$\text{the value of } T\ulcorner\psi\urcorner = \text{the value of } \psi.$$

Pick a reading for connectives, quantifiers, and the truth predicate that can be expressed in this way. If we can identify all the equations that are relevant to the evaluation of the sentences in a loop, and if the resulting system has a unique solution in our value space V, it seems natural to assign the respective values to the members of the loop, since the values we found are the only ones that respect the chosen meaning of the logical constants and T.

Let me explain this point with an example: the liar. Following the method I described informally, in order to assign a value to $\neg Tt_\lambda$, we must consider the value of the non-negated sentence Tt_λ, since $\neg$ is the principal connective of $\neg Tt_\lambda$. We represent graphically this step with a downward arrow from $\neg Tt_\lambda$ to Tt_λ.

$$\begin{array}{c} \neg Tt_\lambda \\ \downarrow \\ Tt_\lambda \end{array}$$

Figure 5.1: From $\neg Tt_\lambda$ to Tt_λ

To determine the value of Tt_λ, however, we must determine the value of $\neg Tt_\lambda$, because we adopt a naïve reading for T and because t_λ evaluates to the code of the

[3] This idea appears in some graph-theoretic semantics, e.g. [Gaifman 1992] and [Gaifman 2000].

entire liar sentence. But then we are in a loop, since we have already met another occurrence of $\neg Tt_\lambda$ in our analysis – it was out starting point. We represent this step with an upward arrow from Tt_λ to such occurrence of $\neg Tt_\lambda$.

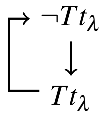


Figure 5.2: Looping graph for the liar sentence

Recognizably, the search is finished: any attempted further analysis would only gain us another turn of the merry-go-round. This loop provides enough information to give a unique value to every sentence in it, using the clauses from above:

$$\begin{cases} \text{the value of } \neg Tt_\lambda & = & 1 - \text{ the value of } Tt_\lambda, \\ \text{the value of } Tt_\lambda & = & \text{the value of } \neg Tt_\lambda. \end{cases}$$

Substituting x for “the value of $\neg Tt_\lambda$” and y for “the value of Tt_λ”, we have a system with a unique solution: $x = 1/2 = y$, assuming that $1/2$ is in the value space V. So, identifying all the relational information needed to evaluate $\neg Tt_\lambda$ and using value-functional equations, we can give a value to the liar sentence (and to Tt_λ).

In some other case, we are not so lucky. Consider the truth-teller, τ. Let’s draw a graph for this sentence, following the intuitive practice that guided us above.

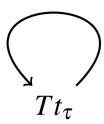


Figure 5.3: Looping graph for the truth-teller sentence

Also in this case the search for sufficient information to evaluate Tt_τ is finished; such information, however, does not allow us to assign a unique numerical value to this sentence. In fact, the only relational information provided by the above loop is the equation associated with Tt_τ, namely "value of $Tt_\tau =$ value of Tt_τ", that has every element of the value space as a solution.

In the next subsection, I will give a more rigorous formal presentation to the kind of analysis and evaluation of $\mathscr{L}_T^{\rightarrow}$-sentences that I just outlined.

5.1.2 Semantic graphs: formal presentation and basic properties

The first thing to do is to make precise the process of "searching for enough information to evaluate a sentence": this is done associating with each $\mathscr{L}_T^{\rightarrow}$-sentence a unique directed labelled rooted graph. In order to do this, let me begin by recalling some graph-theoretic notions, adapted to the present context.

Definition 5.1.1. (Some preliminary graph-theoretic notions)[4]
A *directed graph* is a pair $\langle \mathtt{N}, \mathtt{S} \rangle$ made of a set of *nodes* ($\mathtt{N}$) and a set of *edges* ($\mathtt{S}$).

The edges are ordered pairs of nodes; for simplicity, sometimes I identify a graph with its edges. I will use $\mathtt{v}, \mathtt{w}, \ldots, \mathtt{v_n}, \mathtt{w_n}, \ldots$ to range over nodes. Finally, define the following notions, relative to an arbitrary directed graph $\langle \mathtt{N}, \mathtt{S} \rangle$.

- A directed graph $\langle \mathtt{N}^\dagger, \mathtt{S}^\dagger \rangle$ is a *sub-graph* of a directed graph $\langle \mathtt{N}, \mathtt{S} \rangle$, in symbols $\langle \mathtt{N}^\dagger, \mathtt{S}^\dagger \rangle \subseteq_g \langle \mathtt{N}, \mathtt{S} \rangle$, if $\mathtt{N}^\dagger \subseteq \mathtt{N}$ and $\mathtt{S}^\dagger \subseteq \mathtt{S}$.

- A finite non-empty tuple of edges $\mathtt{P}$ is a *path* if, for every edge $\langle \mathtt{v}, \mathtt{w} \rangle$ in $\mathtt{P}$, $\mathtt{v}$ is either the first element of the first edge of $\mathtt{P}$ or the second element of the edge that precedes $\langle \mathtt{v}, \mathtt{w} \rangle$ in $\mathtt{P}$, and dually $\mathtt{w}$ is either the second node in the last edge of $\mathtt{P}$ or the first element of the edge that follows $\langle \mathtt{v}, \mathtt{w} \rangle$ in $\mathtt{P}$. More intuitively, a path is a finite non-empty tuple of edges of the following form:

$$\Big\langle \langle \mathtt{v_1}, \mathtt{v_2} \rangle, \langle \mathtt{v_2}, \mathtt{v_3} \rangle \ldots, \langle \mathtt{v_{n-2}}, \mathtt{v_{n-1}} \rangle, \langle \mathtt{v_{n-1}}, \mathtt{v_n} \rangle \Big\rangle.^5$$

[4] For comprehensive handbooks of graph theory, see [Bondy and Murty 2008] and [Diestel 2010].

[5] In graph theory, a path within a directed graph is usually defined as a tuple of alternating nodes and edges, that begins and ends with nodes, and where every edge connects the two nodes that

- A path $\mathtt{P}$ is *from* $\mathtt{v}$ *to* $\mathtt{w}$ if $\mathtt{v}$ is the first element of its first pair and $\mathtt{w}$ is the second element of its last pair.

- A finite tuple of nodes $\mathtt{P}'$ is a *sub-path* of a path $\mathtt{P}$, in symbols $\mathtt{P}' \subseteq_p \mathtt{P}$, if $\mathtt{P}' \subseteq \mathtt{P}$ and $\mathtt{P}'$ is a path. The notion of being a proper sub-path, in symbols $\subsetneq_p$, is defined as obvious.

- A path $\mathtt{P}$ in $\langle \mathtt{N}, \mathtt{S} \rangle$ is *maximal* if there is no path $\mathtt{P}'$ in $\langle \mathtt{N}, \mathtt{S} \rangle$ s.t. $\mathtt{P} \subsetneq_p \mathtt{P}'$.

- A path $\mathtt{P}$ is a *loop* if, for every $\mathtt{v} \in Tc(\mathtt{P})$ (the transitive closure of $\mathtt{P}$), there is a path $\mathtt{P}' \subseteq_p \mathtt{P}$ from $\mathtt{v}$ to $\mathtt{v}$ s.t. $Tc(\mathtt{P}') = Tc(\mathtt{P})$ and every node in $\mathtt{P}'$ appears exactly twice in it.[6] A path $\mathtt{P}$ is *straight* if no subpath of $\mathtt{P}$ is a loop.

- Suppose that there is a unique straight path $\mathtt{P}$ from $\mathtt{w}$ to $\mathtt{v}$, for $\mathtt{w}, \mathtt{v} \in \mathtt{N}$. Then, the set $\mathtt{Pred}_{\mathtt{w}}(\mathtt{v}) := Tc(\mathtt{P}$ minus its last pair) is the set of the *predecessors of* $\mathtt{v}$ *from* $\mathtt{w}$. If $\mathtt{v}_1 \in \mathtt{Pred}_{\mathtt{w}}(\mathtt{v})$, then $\mathtt{v}$ is a *successor* of $\mathtt{v}_1$ with respect to $\mathtt{w}$. If $\langle \mathtt{v}_1, \mathtt{v} \rangle$ belongs to $\mathtt{P}$, then $\mathtt{v}$ is an *immediate successor* of $\mathtt{v}_1$ and $\mathtt{v}_1$ is an *immediate predecessor* of $\mathtt{v}$. In the application that follows I will deal with rooted graphs and the parameter $\mathtt{w}$ will always be the root.

- A node $\mathtt{v} \in \mathtt{N}$ is a *simple leaf* if there is no node $\mathtt{w} \in \mathtt{N}$ s.t. $\langle \mathtt{v}, \mathtt{w} \rangle$ in $\mathtt{S}$.

▲

We can now define formally the graphs mentioned above, to be used to unravel and represent the structure of $\mathscr{L}_T^{\rightarrow}$-sentences. For simplicity, assume the existence of countably infinitely many distinct nodes to be used in graphs, and let *en* be a fixed, non-repeating enumeration of all the closed $\mathscr{L}_T^{\rightarrow}$-terms.

precede and follow it. More intuitively, it is an object of the following form:

$$\Big\langle \mathtt{v}_1, \langle \mathtt{v}_1, \mathtt{v}_2 \rangle, \mathtt{v}_2, \langle \mathtt{v}_2, \mathtt{v}_3 \rangle \ldots, \mathtt{v}_{n-2}, \langle \mathtt{v}_{n-2}, \mathtt{v}_{n-1} \rangle, \mathtt{v}_{n-1}, \langle \mathtt{v}_{n-1}, \mathtt{v}_n \rangle, \mathtt{v}_n \Big\rangle.$$

In my definition, I removed the nodes and left only the edges, for notational simplicity.

[6] What I call "loop" is more commonly referred to as "simple cycle" in graph theory.

Definition 5.1.2. (Semantic graphs for $\mathscr{L}_T^{\rightarrow}$-sentences)

For every $\varphi \in \mathscr{L}_T^{\rightarrow}$, the *semantic graph generated by* φ is the directed labelled rooted graph $\langle \mathtt{N}_\varphi, \mathtt{S}_\varphi \rangle$, or simply $\mathtt{S}_\varphi$, whose root is labelled with φ and that is generated by the rules below. In semantic graphs, every node is labelled with exactly one $\mathscr{L}_T^{\rightarrow}$-sentence, so I also define define a *labelling function* l_φ, from $\mathtt{N}_\varphi$ to $\mathscr{L}_T^{\rightarrow}$-sentences, together with each graph.

(G1) There is a node $\mathtt{r}$ s.t. $l_\varphi(\mathtt{r}) = \varphi$ that belongs to $\mathtt{N}_\varphi$. $\mathtt{r}$ is the *root* of $\mathtt{S}_\varphi$. In what follows, I will often write $\mathtt{Pred}_\varphi(\mathtt{v})$ rather than $\mathtt{Pred}_\mathtt{r}(\mathtt{v})$, for φ the sentence labelling the root node of the semantic graph in question.

Let $\mathtt{v} \in \mathtt{N}_\varphi$, $l_\varphi(\mathtt{v}) = \sigma$, and apply the following clauses:

(G2) If σ is an atomic arithmetical sentence, then $\mathtt{v}$ is a simple leaf.

(G3) If σ is $\neg\psi$ and for every $\mathtt{w} \in \mathtt{Pred}_\varphi(\mathtt{v})$ we have that $l_\varphi(\mathtt{w}) \neq \psi$, then add a new node $\mathtt{v}'$ to $\mathtt{N}_\varphi$ and the edge $\langle \mathtt{v}, \mathtt{v}' \rangle$ to $\mathtt{S}_\varphi$, and put $l_\varphi(\mathtt{v}') = \psi$.

(G4) If σ is $\psi \circ \chi$ (where $\circ$ is either $\wedge$, or $\vee$, or $\rightarrow$) and for every $\mathtt{w}_1 \in \mathtt{Pred}_\varphi(\mathtt{v})$ we have that $l_\varphi(\mathtt{w}_1) \neq \psi$, then add a new node $\mathtt{v}'$ to $\mathtt{N}$ and the edge $\langle \mathtt{v}, \mathtt{v}' \rangle$ to $\mathtt{S}_\varphi$, and put $l_\varphi(\mathtt{v}_\mathtt{j}) = \psi$; if for every $\mathtt{w}_2 \in \mathtt{Pred}_\varphi(\mathtt{v})$ we have that $l_\varphi(\mathtt{w}_2) \neq \chi$, then add a new node $\mathtt{v}''$ to $\mathtt{N}_\varphi$ and the edge $\langle \mathtt{v}, \mathtt{v}'' \rangle$ to $\mathtt{S}_\varphi$, and put $l_\varphi(\mathtt{v}'') = \chi$.

(G5) If σ is $\forall x \chi(x)$ then for every $n \in \omega$, if for every $\mathtt{w} \in \mathtt{Pred}_\varphi(\mathtt{v})$ we have that $l_\varphi(\mathtt{w}) \neq \chi(en(n))$ then add a new node $\mathtt{v}_{\mathtt{i(en(n))}}$ to $\mathtt{N}_\varphi$ and the edge $\langle \mathtt{v}, \mathtt{v}_{\mathtt{i(en(n))}} \rangle$ to $\mathtt{S}_\varphi$, and put $l_\varphi(\mathtt{v}_{\mathtt{i(en(n))}}) = \chi(en(n))$.[7]

(G6) If σ is Tt for a $\mathscr{L}_T^{\rightarrow}$-term t, then:

(G6.1) If t denotes the code of a $\mathscr{L}_T^{\rightarrow}$-sentence ψ and for every $\mathtt{w} \in \mathtt{Pred}_\varphi(\mathtt{v})$ we have that $l_\varphi(\mathtt{w}) \neq \psi$, then add a new node $\mathtt{v}'$ to $\mathtt{N}_\varphi$ and the edge $\langle \mathtt{v}, \mathtt{v}' \rangle$ to $\mathtt{S}_\varphi$, and put $l_\varphi(\mathtt{v}') = \psi$.

[7] $\mathtt{i(en(n))}$ is an index uniquely associated with this node: every semantic graph has at most countably infinitely many nodes, so there are many simple functions that can assign suitable indices.

(G6.2) If t does not denote the code of a $\mathscr{L}_T^{\rightarrow}$-sentence, $\mathtt{v}$ is a simple leaf.

$(\mathtt{Loop})$ Whenever, in **(G3)-(G6.2)**, there is a $\mathtt{w} \in \mathtt{Pred}_\varphi(\mathtt{v})$ s.t. $l_\varphi(\mathtt{w}) = \psi$ (or χ, or $\chi(en(n))$, depending on the case), we have that $\langle \mathtt{v}, \mathtt{w} \rangle \in \mathtt{S}_\varphi$.[8]

▲

Lemma 5.1.3.

For every $\varphi \in \mathscr{L}_T^{\rightarrow}$, for every node $\mathtt{v} \in \mathtt{N}_\varphi$, the set $\mathtt{Pred}_\varphi(\mathtt{v})$ is finite.

Proof sketch.

This claim is established by induction. If $\mathtt{v}$ is the root, the claim holds trivially. If $\mathtt{v}$ is added via one of **(G1)-(G6.2)** (the clause $(\mathtt{Loop})$ does not add any new node), assume as IH that the cardinality of $\mathtt{Pred}_\varphi(\mathtt{v})$ is n and note that every node $\mathtt{v}'$ possibly added to the graph by one of **(G1)-(G6.2)** is an immediate successor of $\mathtt{v}$, thus making the cardinality of $\mathtt{Pred}_\varphi(\mathtt{v}')$ equal to $n+1$. □

Theorem 5.1.4.

For every $\varphi \in \mathscr{L}_T^{\rightarrow}$, there is exactly one semantic graph $\langle \mathtt{N}_\varphi, \mathtt{S}_\varphi \rangle$ and exactly one labelling function l_φ.

Proof sketch.

Let φ be any $\mathscr{L}_T^{\rightarrow}$-sentence. Note that the clauses of Definition 5.1.2 yield an inductive simultaneous definition of $\mathtt{N}_\varphi$, $\mathtt{S}_\varphi$, and l_φ (by inspection), with identity and $\mathtt{Pred}_\varphi(\mathtt{v})$ as the only symbols to occur negatively in it ($\mathtt{Pred}_\varphi(\mathtt{v})$ is in the antecedent of some classical conditionals): however, identity is allowed negatively in inductive definitions, and $\mathtt{Pred}_\varphi(\mathtt{v})$ is always finite by Lemma 5.1.3, and hence recursive. Existence and uniqueness of $\langle \mathtt{N}_\varphi, \mathtt{S}_\varphi \rangle$ and l_φ follow. □

[8] The clause $(\mathtt{Loop})$ is in a compressed form in order to keep the whole definition short: spelling it out entirely would require to repeat all the cases in **(G3)-(G6.2)**. Intuitively, $(\mathtt{Loop})$ applies when one is examining a node $\mathtt{v}$ labelled with σ and it is not possible to generate a successor node following the conditions in **(G3)-(G6.2)** because of the presence of a predecessor of $\mathtt{v}$ which is labelled as one of the immediate proper subsentences of σ, or as ψ if σ is $T\ulcorner\psi\urcorner$.

To get an intuitive idea of the effect of the clause (Loop), recall that a $\mathscr{L}_T^{\rightarrow}$-formula φ is *grounded in* $\mathscr{L}$ (or just *grounded*) if, roughly, decomposing it into its subformulae and disquoting every subformula of the form $T\ulcorner\psi\urcorner$ we find in the process, we end up with only formulae of $\mathscr{L}$.[9] (Loop) terminates with a loop some paths generated by otherwise ungrounded sentences. For example, if we constructed the semantic graph of λ using Definition 5.1.2 *without the clause* (Loop), we would get an infinite downward succession of nodes:

$$\begin{array}{c} \neg T t_\lambda \\ \downarrow \\ T t_\lambda \\ \downarrow \\ \neg T t_\lambda \\ \downarrow \\ \vdots \end{array}$$

Figure 5.4: The semantic graph of $\neg T t_\lambda$ built without the clause (Loop)

The difference between Figure 5.4 and Figure 5.2 shows that the conditions for the termination of a path in a semantic graph may be used to expand the notion of groundedness I just recalled with the inclusion of paths ending in loops.

Definition 5.1.5. (Groundedness*)

Let's say that a sentence $\varphi \in \mathscr{L}_T^{\rightarrow}$ is *grounded** if every maximal straight path starting from the root of $\langle \mathtt{N}_\varphi, \mathtt{S}_\varphi \rangle$ is s.t. the second node in its last edge is either a simple leaf or the clause (Loop) applies to it. ▲

Lemma 5.1.6. [10]

Every grounded $\mathscr{L}_T^{\rightarrow}$-sentence is grounded*, but not vice versa.

[9]For an analysis of groundedness in Kripke's theory, see [Yablo 1982] and [Leitgeb 2005]. Following Yablo (at least in part), one could say that a sentence $\varphi \in \mathscr{L}_T^{\rightarrow}$ is *grounded* if constructing its semantic graph using Definition 5.1.2 *without the clause* (Loop), the resulting graph has no infinite path and in every maximal path in the resulting graph the last node is labelled with an atomic arithmetical sentence or a sentence Tt, where t does not denote the code of a $\mathscr{L}_T^{\rightarrow}$-sentence – by Definition 5.1.1 paths are finite tuples, but infinite tuples can be defined via infinite Cartesian products.

[10]The proof is immediate, due to semantic graphs such as Figure 5.2.

5.1.3 Semantic graphs: informal description and examples

Let me gloss informally the construction of semantic graphs. We start with a root node labelled as the sentence generating the graph. Then, we analyze the nodes belonging to the graph in stages, starting from the root, and new labelled nodes and new edges are added to the graph as a result of this analysis. If, e.g., we are analyzing a node $\mathtt{v}$ labelled with a negation $\neg\psi$, we check whether a node labelled with ψ occurs amongst its predecessors: if so, the clause ($\mathtt{Loop}$) applies and an edge between $\mathtt{v}$ and this predecessor is added to the graph. If every predecessor of $\mathtt{v}$ is different from ψ, instead, we add a fresh node $\mathtt{v}'$ to the graph and the edge $\langle \mathtt{v}, \mathtt{v}' \rangle$, and we label $\mathtt{v}'$ as ψ. The other connectives work similarly. If $\mathtt{v}$ is labelled with a universally quantified sentence $\forall x \chi(x)$, we consider all its instances, eliminating the prefix $\forall x$ and replacing the designated variable x with every closed $\mathscr{L}_T^{\rightarrow}$-term. For uniformity, I fixed an enumeration of the closed $\mathscr{L}_T^{\rightarrow}$-term, the function en. For every sentence $\chi(en(n))$, we check whether there is a predecessor of $\mathtt{v}$ labelled as $\chi(en(n))$: if so, ($\mathtt{Loop}$) applies to it as indicated above; if not, we add an immediate successor of $\mathtt{v}$ labelled as $\chi(en(n))$ to the graph – so, every node labelled with a universally quantified sentence has countably infinitely many immediate successors. Sentences of the form Tt, finally, work as negated sentences, with the difference that the truth predicate may apply to a term that does not denote the code of a $\mathscr{L}_T^{\rightarrow}$-sentence, in which case the analysis stops here.

Now I give some examples to clarify how semantic graphs can appear. A simple combination of loops and groundedness is given in the semantic graph of κ:

$$\begin{tikzcd} & Tt_\kappa \to 0 = 1 \arrow[dl] \arrow[dr] & \\ Tt_\kappa \arrow[ur, bend left] & & 0 = 1 \end{tikzcd}$$

Figure 5.5: The semantic graph of Curry's sentence

A node may appear in countably infinitely many loops, as in McGee's sentence.

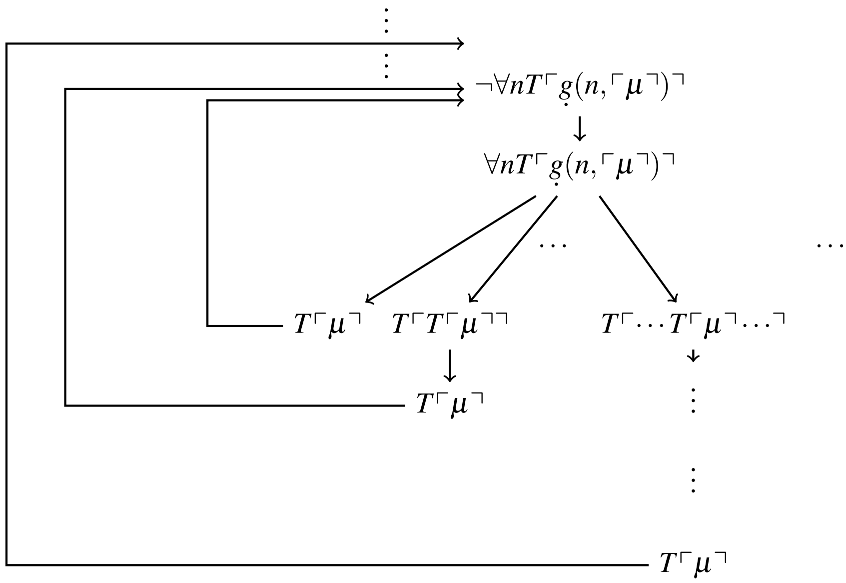


Figure 5.6: The semantic graph of McGee's sentence

μ is grounded*, even if infinitely many paths must be checked to see it. This is not specific to groundedness*: take the semantic graph of, say, $\neg\forall n T\ulcorner \underset{\cdot}{g}(n,\ulcorner 0 = 0\urcorner)\urcorner$ ("not all the iterations of the truth predicate in front of $0 = 0$ are true").

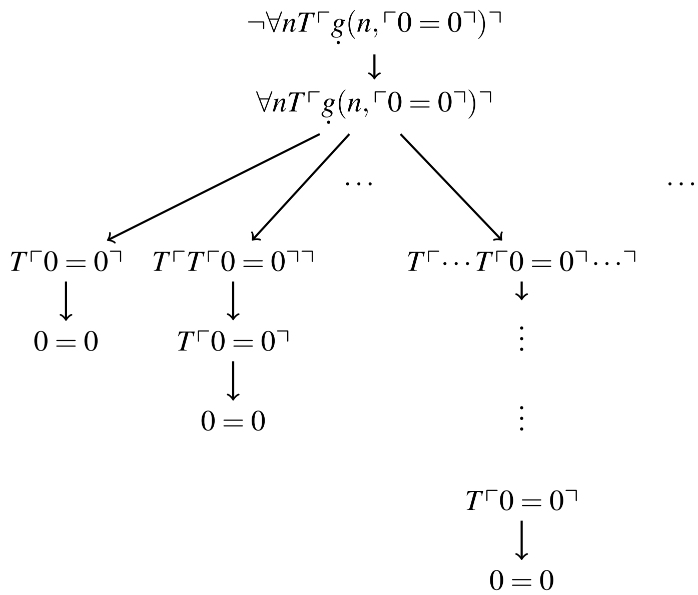


Figure 5.7: The semantic graph of a grounded sentence

As it can also be seen from the figures, the groundedness* of μ is a generalization (with the inclusion of loops) of the groundedness of $\neg\forall n T\ulcorner \underset{\cdot}{g}(n,\ulcorner 0 = 0\urcorner)\urcorner$.

Any Yablo sentence, instead, is ungrounded* and generates a graph that has only infinitely descending paths .[11]

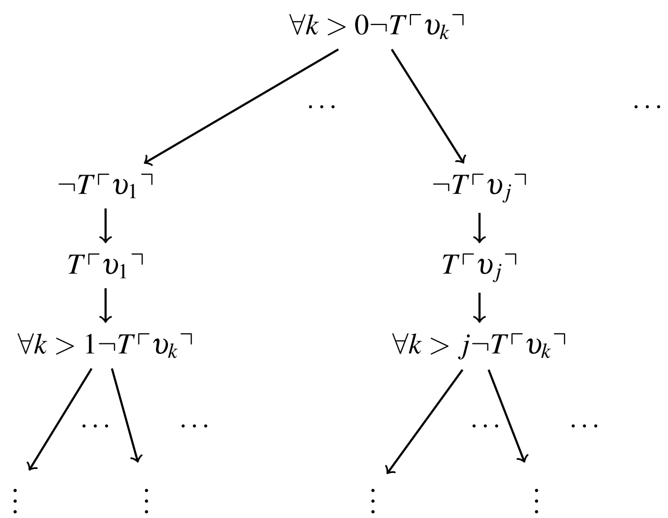


Figure 5.8: The semantic graph of the 0-th Yablo sentence

In the next section, I will use semantic graphs to provide a semantics for $\mathscr{L}_T^{\rightarrow}$-sentences, as sketched in Subsection 5.1.1. I will construct evaluation functions for the *nodes* of each semantic graph – as it was done informally for λ and τ. Using some structural properties of semantic graphs, I will use the evaluations defined on the nodes of semantic graphs to construct a unique evaluation for *sentences*, called the *canonical evaluation*, which is the model I advocate in this chapter.

5.2 Evaluating nodes in semantic graphs

Now I define evaluations to give a value (whenever possible) to nodes of semantic graphs. These evaluations are assignments, in ordinal stages, that go from the bot-

[11] See [Yablo 1993]. υ_0 is the 0-th Yablo sentence, i.e. $\forall k > 0 \neg T\ulcorner \upsilon_k \urcorner$, and υ_k is the k-th Yablo sentence. This formalization is somehow sloppy (e.g., it would be more appropriate to use a satisfaction predicate) but it will suffice for the present purposes.

tom of a graph to its root. Any such evaluation is partial, as some graphs have only infinitely descending paths, as in the case of Yablo sentences. Before describing the evaluations, however, it is fundamental to select an interpretation for the logical vocabulary of $\mathscr{L}_T^{\rightarrow}$ as well as for the truth predicate.

5.2.1 The meaning(s) of connectives and quantifiers

As suggested at the end of Section 4.6, an interesting choice to address FIELD'S PROGRAM is to adopt the Łukasiewicz reading for connectives and quantifiers, even if the combination of this choice and weak naïveté entails that not every sentence can be assigned a numerical value.[12] The Łukasiewicz algebraic reading of connectives and quantifiers was given in Definition 4.1.1; let me recall, however, that the Łukasiewicz interpretation of $\neg$, $\wedge$, $\vee$, and $\forall$ is essentially that of Kripke's theory,[13] while for the conditional we have this clause:

$$\text{the value of } \varphi \rightarrow \psi = \min[1, (1 - \text{ the value of } \varphi + \text{ the value of } \psi)].$$

A Łukasiewicz numerical value space V is either any set $\{0, {}^{1}/_{n+1}, \ldots, {}^{n}/_{n+1}, 1\}$, for n an odd positive integer, or the real interval $[0,1]$. The Łukasiewicz algebraic reading of connectives and quantifiers does not change with $|V|$, the cardinality of V. Also, the semantics I will give applies uniformly to every choice of V. So, in the following, V will indicate a fixed, arbitrary Łukasiewicz value space.

Łukasiewicz semantics confers an interesting reading on the conditional, that can be used to express a comparison between the semantic values of antecedent and consequent, in the sense explored in Subsection 3.5.1 and applied to Łukasiewicz 3-valued conditional in Subsection 4.4.2. It is easy to realize that this reading extends to the conditional of *every* Łukasiewicz semantics. In particular, in every

[12] See Theorem 4.1.2.
[13] See Definition 1.4.4 and Lemma 2.4.3.

Łukasiewicz evaluation, $\varphi \to \psi$ has value 1 if the value of φ is less than or equal to the value of ψ, and it has a value that measures how close the difference of the values of φ and ψ comes to a 1-valued conditional, if the value of φ is strictly greater than the value of ψ. This reading is philosophically significant, for the reasons given in Subsection 3.5.1. Let me stress, however, that the model I will give is, in its main properties, independent from this interpretation of connectives and quantifiers and can accommodate different readings for them.

5.2.2 Evaluating nodes in semantic graphs: some heuristics

I will now give an informal sketch of the evaluations of nodes in semantic graphs, in order to facilitate the reading of the technical presentation given in Subsection 5.2.3. Let's begin with the treatment of loops.

How to evaluate nodes in loops

Nodes that are not in loops do not pose particular problems, so let's see how to deal with loops. We have seen how to treat λ: would the very same procedure give a unique value to, e.g., Curry's sentence? Figure 5.5 gives the following system (I replaced "$0 = 1$" in Curry's sentence with "$\bot$", to avoid notational confusions with the numerical values of V):

$$\begin{cases} \text{value of } Tt_\kappa \to \bot = \min[0, (1 - \text{ value of } Tt_\kappa + \text{ value of } \bot)] \\ \text{value of } Tt_\kappa = \text{value of } Tt_\kappa \to \bot. \end{cases}$$

Substituting variables for the value of the nodes in the loop, we get:

$$\begin{cases} x = \min[1, (1 - y + \text{ value of } \bot)] \\ y = x. \end{cases}$$

"Value of $\perp$" is not replaced by a variable because $\perp$ does not belong to the loop. Now, the value of $\perp$ is 0. *After this assignment*, we update the previous system to:

$$\begin{cases} x = \min[1,(1-y+0)] \\ y = x, \end{cases}$$

which has a unique solution, namely $x = 1/2 = y$. In the case of λ the loop alone determines the values of all its elements, in the case of κ it does not.[14]

The node labelled with $\perp$ has to be evaluated in order to assign a value to the nodes in the loop not only because the system has no solution otherwise: it is a safe practice that may avoid serious troubles. Suppose that $\mathtt{v}$ belongs to a loop $\mathtt{L}_1$, that $\mathtt{v}$ has $\mathtt{v}_1$ and $\mathtt{v}_2$ as immediate successors, and that $\mathtt{v}_1$ is in the loop $\mathtt{L}_1$, but $\mathtt{v}_2$ is not. If we could associate an equation system E_1 with the nodes in $\mathtt{L}_1$ alone and E_1 had a unique solution, independent from the value of $\mathtt{v}_2$, we should not try to solve it (yet): some successor of $\mathtt{v}_2$, in fact, may loop back to $\mathtt{v}$, and the system E_1 we considered at the beginning would not be a complete representation of all the value relations between the sentences labelling the nodes in $\mathtt{L}_1$. E_1 would be missing at least one equation (in at least one of the unknowns of E_1), which should be included to evaluate the nodes in $\mathtt{L}_1$. But adding an equation to E_1 would produce a new system E_2 that may have no solutions.

Also, equations are not always needed to evaluate nodes in loops. Suppose that $\mathtt{L}_2$ is a loop, that $\mathtt{v}$ is one of the nodes in $\mathtt{L}_2$, and that $\mathtt{v}$ is labelled as $\sigma_1 \to 0 = 0$, with the node labelled as σ_1 in $\mathtt{L}_2$ but not the node labelled as $0 = 0$. The value of $\mathtt{v}$ does not depend on the semantic relations it bears to other nodes in $\mathtt{L}_2$: it has value 1 because of the value of $0 = 0$. It is clear that, in the strong Kleene paradigm, the

[14]It does not matter that, in the case of λ, the loop exhausts $\mathtt{S}_\lambda$. Consider the case of $\lambda \to \lambda$: the resulting graph does not present a unique loop, but rather two loops and a node, the root, not involved in any of them. Also in the case of $\lambda \to \lambda$, however, the loops alone suffice to determine the value of all the nodes in them – such values are then used to evaluate the root.

value of a conditional with a 1-valued consequent is 1, independently of any value relation that its antecedent may bear to other sentences.

Some preliminary notions

In order to proceed to the construction of the evaluation functions, we need to distinguish some types of nodes, and to fix our stock of equations.

Definition 5.2.1. (Types of nodes in a semantic graph)

For every $\varphi \in \mathscr{L}_T^{\rightarrow}$, a node $\mathtt{v} \in \mathtt{N}_\varphi$ is:

- a *simple leaf*, as per Definition 5.1.1.
- a *looping leaf*, if $(\mathtt{Loop})$ is the only clause in Definition 5.1.2 that applies to $\mathtt{v}$ in the construction of $\langle \mathtt{N}_\varphi, \mathtt{S}_\varphi \rangle$.
- a *simple point*, if it is not a leaf and it is not in a loop.
- a *looping point*, if it is not a leaf and it is in a loop and its immediate predecessor is in a loop.
- a *loop top*, if it is not a leaf and it is in a loop and it is the root or its immediate predecessor is a simple point. If $\mathtt{w}$ is in a loop, $\mathtt{w}'$ is *the loop top of* $\mathtt{w}$ if $\mathtt{w}'$ is a loop top, it is in $\mathtt{Pred}_\varphi(\mathtt{w})$, and there is a straight path $\mathtt{P}$ from $\mathtt{w}'$ to $\mathtt{w}$ s.t. $\mathtt{w}'$ is the only loop top in $\mathtt{P}$.

The above notions are well-defined, and every looping point has one loop top. ▲

Definition 5.2.2. (Łukasiewicz equations for $\neg$, $\wedge$, $\rightarrow$, $\forall$, and a naïve T)

I define a language $\mathscr{L}_U$ to represent conveniently the equations definable from the value functions of Łukasiewicz logics, plus naïve truth. $\mathscr{L}_U$ is made of:

- Countably infinitely many new variables $VAR_U := \{u_{\varphi_1}, \ldots, u_{\varphi_n}, \ldots\}$, where φ_n is the n-th element of a non-repeating enumeration of $\mathscr{L}_T^{\rightarrow}$-sentences.

- A set of constants CON_U s.t. it contains exactly one term for each element of the value space V. I will use $c_{v_1}, \ldots, c_{v_n}, \ldots$ to range over CON_U. Note that, if $V = [0,1]$, CON_U is uncountable. However, this has little consequences, as CON_U is used to reproduce the elements of V that are unique solutions of equation systems generated by semantic graphs, but semantic graphs generate only systems with countably many equations, variables, and unique solutions. So, I will stick to the above (countable) notation. I do not put $CON_U = V$ to avoid confusion with the numerical value space, although the two sets will be identified in practice.

- Let $s_1, \ldots, s_m, s_n, \ldots$ be $\mathscr{L}_U$-terms. Then, $id(s_m)$ (where $id(x)$ is the identity function), $c_1 - s_m$, $\min(s_m, s_n)$, $\min[c_1, (c_1 - s_m + s_n)]$, $\inf\{s_1, \ldots, s_m, s_n, \ldots\}$ are $\mathscr{L}_U$-terms (and nothing else is). Let $h_i(\ldots)$ range over the term-forming $\mathscr{L}_U$-functions.

- Let s_m, s_n be $\mathscr{L}_U$-term. Then $\mathbf{s_m =_U s_n}$ is an atomic $\mathscr{L}_U$-formula (and nothing else is). Call this set Ł, and its element *Łukasiewicz equations*.

▲

Informal description of the evaluations of nodes in semantic graphs

Since I aim at evaluating nodes within semantic graphs, I define one evaluation function per graph, parametrized to the sentence labelling the root. The evaluation function associated with the semantic graph generated by φ will be denoted by e_φ. The evaluations send nodes to Ł $\cup V$, starting from leaves and "climbing up" toward the root, marking this progression with ordinals. The range of the evaluations is Ł $\cup V$, since in order to assign a value in V we may have to pass through a stage at which we give to a node an equation in Ł. Moreover, as we have seen, some equation systems have no solutions or have multiple solutions, in which case we leave Łukasiewicz equations associated with the corresponding nodes.

At stage $\overline{0}$, all the leaves are evaluated: simple leaves are assigned a classical value and looping leaves are associated with Łukasiewicz equations, to be solved later. Non-leaves are analyzed at non-$\overline{0}$ stages. I explain the non-leaf case with an example: suppose we are evaluating v, not a leaf, labelled as $\psi \wedge \chi$, at level α. v has at least one immediate successor, since it is not a leaf, labelled as ψ or χ.

- If at some level $\beta < \alpha$, one of the successors has value 0, than v has value 0 ("a conjunction with a 0-valued conjunct is false").

- If the node has two successors and both have a value in V at some levels $\beta, \gamma < \alpha$, than v has the minimum of their values from α onwards.

- Equations are assigned if values in V cannot be given. We give v its equation in Ł at α if (1) v has two successors, both are given an equation or a value at $\beta, \gamma < \alpha$, at least one equation, no value 0 is obtained, or (2) v loops back to a predecessor and has one successor that already has an equation or a value different from 0 (this avoids clashes with the two previous bullet points).

- After assigning equations, we want to solve them in V, if possible. A collection of equations is fully examined when we reach its loop top. If v is a loop top, it gets value k at stage α if v had an equation in Ł at some $\beta < \alpha$, the resulting system has only one solution and k is the solution for $\psi \wedge \chi$. If v is not a loop top, it gets value k at stage α if its loop top has a value in V at some $\beta < \alpha$ and v was assigned an equation in Ł at some $\gamma < \beta$, and k is the solution for $\psi \wedge \chi$.

Clauses such as these apply iteratively to every node for which the inductive defining conditions are satisfied – of course we have different clauses for other connectives and quantifiers, and for truth. I stipulate that before starting the evaluation of $\mathtt{N}_{\varphi}$, it is determined what type of node is $\mathtt{v}$, according to Definition 5.2.1.

5.2.3 Evaluating nodes in semantic graphs: formal definition

Definition 5.2.3. (Evaluation functions for nodes in a semantic graph)
For every $\varphi \in \mathscr{L}_T^{\rightarrow}$, define a partial function $e_{\varphi} : Ord \times \mathtt{N}_{\varphi} \rightharpoonup (\text{Ł} \cup V)$ as follows.[15] Suppose one is examining a node $\mathtt{v_n}$, with $l_{\varphi}(\mathtt{v_n}) = \sigma$. There are two main cases.

(I) Suppose $\mathtt{v_n}$ is a leaf. There are only two sub-cases:

(a) By construction, there are only two possibilities.

- σ is an atomic arithmetical sentence. Then for all ordinals α, $e_{\varphi}(\alpha, \mathtt{v_n}) = 1$ if $\mathbb{N} \models \sigma$, and 0 if $\mathbb{N} \not\models \sigma$.
- σ is Tt, t does not denote the code of a $\mathscr{L}_T^{\rightarrow}$-sentence. Then for all ordinals α, $e_{\varphi}(\alpha, \mathtt{v_n}) = 0$.

(b) $\mathtt{v_n}$ is a looping leaf, looping back to nodes $\mathtt{v_1}, \dots, \mathtt{v_p}$ labelled with sentences $\sigma_1, \dots, \sigma_p$.[16] Put:

- $e_{\varphi}(\overline{0}, \mathtt{v_n}) = (\mathbf{u}_{\sigma} =_{\mathbf{U}} \mathbf{h_i}(\mathbf{u}_{\sigma_1}, \dots, \mathbf{u}_{\sigma_\mathbf{p}}))$, where $u_{\sigma_1}, \dots, u_{\sigma_p}$ are the variables associated with $\sigma_1, \dots, \sigma_p$, and h_j is the appropriate term-forming function (depending on the logical form of σ).
- $e_{\varphi}(\alpha, \mathtt{v_n}) = k$, if for every node $\mathtt{v_m} \in \{\mathtt{v_1}, \dots, \mathtt{v_p}\}$ there is an ordinal $\beta < \alpha$ and a $j \in V$ s.t. $e_{\varphi}(\beta, \mathtt{v_m}) = j$ and k is the unique solution for u_{σ}. See the general case below for details of solving Łukasiewicz equations.

(II) $\mathtt{v_n}$ is not a leaf. In what follows, $\alpha > \overline{0}$. I will do only one sub-case in detail.

[15]For simplicity, I will identify $\mathscr{L}_U$-constants and elements of V, writing k for c_k.
[16]By Lemma 5.1.3, a node can loop back to finitely many nodes at most.

(1) σ is $\psi \wedge \chi$. Since $\mathtt{v_n}$ is not a leaf, it has a successor $\mathtt{v_p}$, labelled with ψ, or a successor $\mathtt{v_q}$, labelled with χ, or both. Then:

(1.1) $e_\varphi(\alpha, \mathtt{v_n}) = 0$, if there is an ordinal $\beta < \alpha$ s.t. $e_\varphi(\beta, \mathtt{v_p}) = 0$ or $e_\varphi(\beta, \mathtt{v_q}) = 0$; or

(1.2) $e_\varphi(\alpha, \mathtt{v_n}) = j$, if $\mathtt{v_n}$ has both nodes $\mathtt{v_p}$ and $\mathtt{v_q}$ as successors, and there are ordinals $\beta, \gamma < \alpha$ s.t. $e_\varphi(\beta, \mathtt{v_p}) = j$, $e_\varphi(\gamma, \mathtt{v_q}) = k$ (for $j, k \in V$), and $j \leq k$ (and it yields value k if $k \leq j$); or

(1.3) $e_\varphi(\alpha, \mathtt{v_n}) = (\mathbf{u}_\sigma =_{\mathbf{U}} \min(\mathbf{s_p}, \mathbf{s_q}))$, if $\mathtt{v_n}$ has both $\mathtt{v_p}$ and $\mathtt{v_q}$ as successors and there are ordinals $\beta, \gamma < \alpha$ s.t.

(1.3.1) $e_\varphi(\beta, \mathtt{v_p}), e_\varphi(\gamma, \mathtt{v_q}) \in Ł \cup V$, and

(1.3.2) $e_\varphi(\beta, \mathtt{v_p}) \neq 0$ and $e_\varphi(\gamma, \mathtt{v_q}) \neq 0$, and

(1.3.3) $e_\varphi(\beta, \mathtt{v_p}) \in Ł$ or $e_\varphi(\gamma, \mathtt{v_q}) \in Ł$, and

(1.3.4) s_p is either the constant c_v corresponding to the value in V assigned to $\mathtt{v_p}$ by $e_\varphi(\beta, \mathtt{v_p})$, or the $\mathscr{L}_U$-variable associated with $\mathtt{v_p}$ by $e_\varphi(\beta, \mathtt{v_p})$, and dually for s_q, $\mathtt{v_q}$, and $e_\varphi(\gamma, \mathtt{v_q})$,

or $\mathtt{v_n}$ loops back to a predecessor labelled with ψ, and s_p is the $\mathscr{L}_U$-variable associated with it, and there is an ordinal $\gamma < \alpha$ s.t. $e_\varphi(\gamma, \mathtt{v_q}) \in Ł \cup V$ and $e_\varphi(\gamma, \mathtt{v_q}) \neq 0$ (or dually, reversing $\mathtt{v_p}$ with $\mathtt{v_q}$ and ψ with χ); or

(1.4) $e_\varphi(\alpha, \mathtt{v_n}) = k$, if $\mathtt{v_n}$ is a loop top and there is an ordinal $\beta < \alpha$ s.t. $e_\varphi(\beta, \mathtt{v_n}) \in Ł$, its equation system has a unique solution, and k is the solution for u_σ; or[17]

(1.5) $e_\varphi(\alpha, \mathtt{v_n}) = k$, if $\mathtt{v_n}$ is a looping point, and there is an ordinal $\beta < \alpha$ s.t. the loop top of $\mathtt{v_n}$ has a value in V at β, and there is an

[17] By "its equation system", I mean the equations assigned (at ordinals less than α) to the nodes of which $\mathtt{v_n}$ is the loop top. I do not spell out this specification in detail within the present definition for the sake of readability.

ordinal $\gamma < \beta$ s.t. $e_\varphi(\gamma, \mathrm{v_n}) \in$ Ł, its system has a unique solution, and k is the solution for u_σ.

For brevity, I will only indicate the changes to be made to item (1).

(2) σ is $\psi \vee \chi$. $\mathrm{v_n}$ has a successor $\mathrm{v_p}$ and $l_\varphi(\mathrm{v_p}) = \psi$, or $\mathrm{v_q}$ and $l_\varphi(\mathrm{v_q}) = \chi$, or both.

(2.1) $e_\varphi(\alpha, \mathrm{v_n}) = 1$, if there is an ordinal $\beta < \alpha$ s.t. $e_\varphi(\beta, \mathrm{v_p}) = 1$ or $e_\varphi(\beta, \mathrm{v_q}) = 1$.

(2.2) $e_\varphi(\alpha, \mathrm{v_n}) = k$, if $\mathrm{v_n}$ has $\mathrm{v_p}$ and $\mathrm{v_q}$ as successors and there are ordinals $\beta, \gamma < \alpha$ s.t. $e_\varphi(\beta, \mathrm{v_p}) = j$, $e_\varphi(\gamma, \mathrm{v_q}) = k$ (for $j, k \in V$), and $j \leq k$ (and it yields value j if $k \leq j$).

(2.3) Replace $(\mathbf{u}_\sigma =_\mathbf{U} \min(\mathbf{s_p}, \mathbf{s_q}))$ with $(\mathbf{u}_\sigma =_\mathbf{U} \max(\mathbf{s_p}, \mathbf{s_q}))$ in (1.3) and every occurrence of 0 with 1 in (1.3.2).

(3) σ is $\psi \rightarrow \chi$. $\mathrm{v_n}$ has a successor $\mathrm{v_p}$ and $l_\varphi(\mathrm{v_p}) = \psi$, or $\mathrm{v_q}$ and $l_\varphi(\mathrm{v_q}) = \chi$, or both.

(3.1) $e_\varphi(\alpha, \mathrm{v_n}) = 1$, if there is an ordinal $\beta < \alpha$ s.t. $e_\varphi(\beta, \mathrm{v_p}) = 0$ or $e_\varphi(\beta, \mathrm{v_q}) = 1$.

(3.2) $e_\varphi(\alpha, \mathrm{v_n}) = i$, if $\mathrm{v_n}$ has $\mathrm{v_p}$ and $\mathrm{v_q}$ as successors and there are ordinals $\beta, \gamma < \alpha$ s.t. $e_\varphi(\beta, \mathrm{v_p}) = j$, $e_\varphi(\gamma, \mathrm{v_q}) = k$ (for $j, k \in V$), and $i = \min[1, (1 - j + k)]$.

(3.3) Replace $(\mathbf{u}_\sigma =_\mathbf{U} \min(\mathbf{s_p}, \mathbf{s_q}))$ with $(\mathbf{u}_\sigma =_\mathbf{U} \min[\mathbf{1}, (\mathbf{1} - \mathbf{s_p} + \mathbf{s_q})])$ in (1.3) and the second occurrence of 0 with 1 in (1.3.2).

(4) σ is $\neg\psi$. $\mathrm{v_n}$ has a successor $\mathrm{v_p}$ labelled with ψ.

(4.1) Replace (1.1) and (1.2) with this clause: $e_\varphi(\alpha, \mathrm{v_n}) = j$, if there is an ordinal $\beta < \alpha$ s.t. $e_\varphi(\beta, \mathrm{v_p}) = k$ and $j = 1 - k$.

(4.2) Put the equation $(\mathbf{u}_\sigma =_\mathbf{U} \mathbf{1} - \mathbf{s_p})$ in (1.3), erasing (1.3.2).

(5) σ is $T\ulcorner\psi\urcorner$. $\mathtt{v_n}$ has a successor $\mathtt{v_p}$ labelled with ψ.

(5.1) Replace (4.1) with the following clause: $e_\varphi(\alpha, \mathtt{v_n}) = i$, if there is a $\beta < \alpha$ s.t. $e_\varphi(\beta, \mathtt{v_p}) = i$.

(5.2) Put the equation $(\mathbf{u}_\sigma =_\mathbf{U} \mathbf{id}(\mathbf{s_p}))$ in (1.3), erase (1.3.2).

(6) σ is $\forall x\chi(x)$. $\mathtt{v_n}$ has countably infinitely many successors, as $(\mathtt{Loop})$ applies at most to finitely many predecessors (see Lemma 5.1.3, and Definition 5.1.2 for notation).

(6.1) $e_\varphi(\alpha, \mathtt{v_n}) = 0$, if there is a node $\mathtt{v_{i(en(k))}}$ (successor of $\mathtt{v_n}$) and an ordinal $\beta < \alpha$ s.t. $e_\varphi(\beta, \mathtt{v_{i(en(k))}}) = 0$.

(6.2) $e_\varphi(\alpha, \mathtt{v_n}) = j$, if for all k, $\mathtt{v_{i(en(k))}}$ is a successor of $\mathtt{v_n}$ and there is an ordinal $\beta_{i(en(k))} < \alpha$ s.t. $e_\varphi(\beta_{i(en(k))}, \mathtt{v_{i(en(k))}}) = i$ (for $i \in V$), and $\inf\{e_\varphi(\beta_{i(en(k))}, \mathtt{v_{i(en(k))}}) | k \in \omega\} = j$.

(6.3) Put the equation $(\mathbf{u}_\sigma =_\mathbf{U} \inf\{\mathbf{s_{i(en(k))}} | k \in \omega\})$ in (1.3), with the following conditions.

(6.3.1) For every $k \in \omega$, there is an ordinal $\beta_{i(en(k))} < \alpha$ such that $e_\varphi(\beta_{i(en(k))}, \mathtt{v_{i(en(k))}})$ has a value, or $(\mathtt{Loop})$ applies to $\mathtt{v_n}$ and a predecessor labelled as $\chi(t_i)$ for a closed $\mathscr{L}_T^{\rightarrow}$-term t_i, and $u_{\chi(t_i)}$ is the $\mathscr{L}_U$-variable assigned uniquely to it, and

(6.3.2) these values are always different from 0, and

(6.3.3) some of these values are equations of Ł, and

(6.3.4) $s_{i(en(k))}$ is j if $e_\varphi(\beta_{i(en(k))}, \mathtt{v_{i(en(k))}}) = j$, or $s_{i(en(k))}$ is the $\mathscr{L}_U$-variable associated with $\mathtt{v_{i(en(k))}}$ in $e_\varphi(\beta_{i(en(k))}, \mathtt{v_{i(en(k))}})$.

▲

Theorem 5.2.4. (Basic facts on the parametrized evaluations of nodes)

For every $\varphi \in \mathscr{L}_T^{\rightarrow}$, the following facts hold:

1. Exactly one e_φ exists.

2. The definition of e_φ *reaches a* fixed point *at* ω_1^{CK}.

3. For every $\mathtt{v_n} \in \mathtt{S}_\varphi$*, there is* at most one $vl \in \text{Ł} \cup V$ *s.t.* $e_\varphi(\omega_1^{\text{CK}}, \mathtt{v_n}) = vl$ *(in other words,* e_φ*, seen as an inductively defined relation, is a function).*

4. If $e_\varphi(\alpha, \mathtt{v_n}) = k$*, for* $k \in V$*, then for all ordinals* $\alpha' \geq \alpha$*,* $e_\varphi(\alpha', \mathtt{v_n}) = k$*.*

Proof sketch.

1. From Definition 5.2.3 it is clear that, for every $\mathscr{L}_T^{\rightarrow}$-sentence φ, the function e_φ can be obtained via an inductive definition, that is inductive in: $\langle \mathtt{N}_\varphi, \mathtt{S}_\varphi \rangle$, the set of the simple leaves of $\langle \mathtt{N}_\varphi, \mathtt{S}_\varphi \rangle$, its looping leaves, its loop tops, the nodes of each of its loops tops. All these sets are well-defined (see Definition 5.2.1). Existence and uniqueness of every function e_φ are thus immediate.

2. By item 1 and the fact that ω_1^{CK} is the closure ordinal for inductive definitions over the standard model of arithmetic.[18]

3. By induction on ordinals and our condition on unique solutions for equation systems.

4. By the previous item.

□

Let me show the functioning of the evaluation functions via some examples.

Example 5.2.5. (The liar sentence)
See Figure 5.2. Let $\mathtt{v_1}$ be the root (so $l_\lambda(\mathtt{v_1}) = \neg T t_\lambda$) and $\mathtt{v_2}$ be the node that loops back to $\mathtt{v_1}$ (so $l_\lambda(\mathtt{v_2}) = T t_\lambda$). By Definition 5.2.3, we get:

[18]See, e.g., [Pohlers 2009], pp. 87 and following.

$$e_\lambda(\overline{0}, \mathtt{v}_2) = (\mathbf{u}_{\mathbf{Tt}_\lambda} =_{\mathbf{U}} \mathbf{u}_{\neg\mathbf{Tt}_\lambda}),$$

$$e_\lambda(\overline{1}, \mathtt{v}_1) = (\mathbf{u}_{\neg\mathbf{Tt}_\lambda} =_{\mathbf{U}} \mathbf{1} - \mathbf{u}_{\mathbf{Tt}_\lambda}),$$

$$e_\lambda(\overline{2}, \mathtt{v}_1) = 1/2,$$

$$e_\lambda(\overline{3}, \mathtt{v}_2) = 1/2. \qquad \blacktriangle$$

Example 5.2.6. (Curry's sentence)

See Figure 5.5. Let $\mathtt{v}_1$ be the root, i.e. $l_\kappa(\mathtt{v}_1) = Tt_\kappa \to \perp$, let $\mathtt{v}_2$ be the node labelled with Tt_κ that loops back to $\mathtt{v}_1$, and let $\mathtt{v}_3$ be the node labelled with $\perp$.

$$e_\kappa(\overline{0}, \mathtt{v}_3) = 0,$$

$$e_\kappa(\overline{0}, \mathtt{v}_2) = (\mathbf{u}_{\mathbf{Tt}_\kappa} =_{\mathbf{U}} \mathbf{u}_{\mathbf{Tt}_\kappa \to \perp}),$$

$$e_\kappa(\overline{1}, \mathtt{v}_1) = (\mathbf{u}_{\mathbf{Tt}_\kappa \to \perp} =_{\mathbf{U}} \mathbf{min}[\mathbf{1}, (\mathbf{1} - \mathbf{u}_{\mathbf{Tt}_\kappa} + \mathbf{0})]),$$

$$e_\kappa(\overline{2}, \mathtt{v}_1) = 1/2,$$

$$e_\kappa(\overline{3}, \mathtt{v}_2) = 1/2. \qquad \blacktriangle$$

Example 5.2.7. (Iterated Curry-like sentences)

Let $\kappa^\star$ be $Tt_{\kappa^\star} \to (Tt_\kappa \to \perp)$ where $t_{\kappa^\star} = \ulcorner Tt_{\kappa^\star} \to (Tt_\kappa \to \perp)\urcorner$. I do not include the graph since its structure is very simple, and it is quite similar to $\mathtt{S}_\kappa$. Let $\mathtt{v}_1$ be the root, i.e. $l_{\kappa^\star}(\mathtt{v}_1) = Tt_{\kappa^\star} \to (Tt_\kappa \to \perp)$, $\mathtt{v}_2$ be the node labelled with $Tt_{\kappa^\star}$ that loops back to $\mathtt{v}_1$, $\mathtt{v}_3$ be labelled with $Tt_\kappa \to \perp$, $\mathtt{v}_4$ be the node labelled with Tt_κ that loops back to $\mathtt{v}_3$, and let $\mathtt{v}_5$ be labelled with $\perp$. We have:

$$e_{\kappa^\star}(\overline{0}, \mathtt{v}_5) = 0,$$

$$e_{\kappa^\star}(\overline{0}, \mathtt{v}_4) = (\mathbf{u}_{\mathbf{Tt}_\kappa} =_{\mathbf{U}} \mathbf{u}_{\mathbf{Tt}_\kappa \to \perp}),$$

$$e_{\kappa^\star}(\overline{0}, \mathtt{v}_3) = (\mathbf{u}_{\mathbf{Tt}_{\kappa^\star}} =_{\mathbf{U}} \mathbf{u}_{\mathbf{Tt}_{\kappa^\star} \to (\mathbf{Tt}_\kappa \to \perp)}),$$

$$e_{\kappa^\star}(\overline{1}, \mathtt{v}_3) = (\mathbf{u}_{\mathbf{Tt}_\kappa \to \perp} =_{\mathbf{U}} \mathbf{min}[\mathbf{1}, (\mathbf{1} - \mathbf{u}_{\mathbf{Tt}_\kappa} + \mathbf{0})]),$$

$$e_{\kappa^\star}(\overline{2}, \mathtt{v}_1) = (\mathbf{u}_{\mathbf{Tt}_{\kappa^\star} \to (\mathbf{Tt}_\kappa \to \perp)} =_{\mathbf{U}} \mathbf{min}[\mathbf{1}, (\mathbf{1} - \mathbf{u}_{\mathbf{Tt}_{\kappa^\star}} + \mathbf{u}_{\mathbf{Tt}_\kappa \to \perp})]),$$

$e_{\kappa^\star}(\overline{3}, \mathtt{v}_1) = 3/4$, if $3/4 \in V$; otherwise, $\mathtt{v}_1$ retains its Łukasiewicz equation,

$e_{\kappa^\star}(\overline{4}, \mathtt{v}_2) = 3/4$, if $3/4 \in V$; otherwise, $\mathtt{v}_2$ retains its Łukasiewicz equation,

$$e_{\kappa^\star}(\overline{5}, \mathtt{v}_3) = 1/2,$$

$e_{\kappa^\star}(\overline{6}, \mathtt{v}_4) = 1/2$.[19] ▲

Example 5.2.8. (The truth-teller)

See Figure 5.3. Let $\mathtt{v}_1$ be the only node, labelled as Tt_τ, which loops back to itself.

$$e_\tau(\overline{0}, \mathtt{v}_1) = (\mathbf{u}_{\mathbf{Tt}_\tau} =_{\mathbf{U}} \mathbf{u}_{\mathbf{Tt}_\tau}),$$

$$\vdots$$

$$e_\tau(\omega_1^{\mathrm{CK}}, \mathtt{v}_1) = (\mathbf{u}_{\mathbf{Tt}_\tau} =_{\mathbf{U}} \mathbf{u}_{\mathbf{Tt}_\tau}).$$ ▲

Example 5.2.9. (McGee's sentence)

See figure 5.6. The entire graph is evaluated as a single system of equations, checked for solutions at level $\omega + 1$. The values of all the sentences of the form $T\ulcorner \cdots T\ulcorner \gamma \urcorner \cdots \urcorner$ are equal between them, the node labelled as $\forall n T \ulcorner \underset{\cdot}{g}(n, \ulcorner \gamma \urcorner) \urcorner$ is the infimum of all their values, and it has a value identical to the value of its negation. The only solution for this system is to assign value $1/2$ to every node. ▲

5.3 Evaluating sentences via semantic graphs

As mentioned at the end of Subsection 5.1.1, the evaluations defined on nodes of semantic graphs do not suffice to provide an interpretation of $\mathscr{L}_T^{\to}$-sentences. In

[19] Note that all the iterated Curry-like sentences that are considered in [Field 2008] (pp. 85-86) and that are essentially analogous to $\kappa^\star$ are assigned a numerical value by the evaluation function generated by their semantic graph if V is large enough, and they are assigned a Łukasiewicz equation otherwise. See also footnote 12 in Subsection 4.2.1.

particular, the fact that a node $\mathtt{v}$, labelled with φ and occurring in a semantic graph $\langle \mathtt{N}_\psi, \mathtt{S}_\psi \rangle$, is given a value (at the closure ordinal) by e_ψ does not guarantee that any other node labelled with φ, occurring in any other semantic graph, is given the same value by the respective evaluation functions. To show that this is indeed the case, we must investigate more in depth the structure of semantic graphs. This is the task of the next subsection. In particular, I will prove a structural result on semantic graphs (Theorem 5.3.4), from which the desired uniformity of the evaluation functions defined on nodes follows (Theorem 5.3.7).

5.3.1 Isomorphisms between semantic graphs

The main reasoning of this subsection can be summarized as follows.

- For every $\varphi \in \mathscr{L}_T^{\rightarrow}$, in order to evaluate a node $\mathtt{v}$ of $\mathtt{N}_\varphi$, possibly not all of $\langle \mathtt{N}_\varphi, \mathtt{S}_\varphi \rangle$ is relevant. It is immediate from Definitions 5.1.2 and 5.2.3 that only the sub-graph of $\langle \mathtt{N}_\varphi, \mathtt{S}_\varphi \rangle$ made of the nodes *reachable* in finitely many steps from $\mathtt{v}$ following the arrows (edges) in $\mathtt{S}_\varphi$ is used in building $e_\varphi(\alpha, \mathtt{v})$.

- For every $\varphi \in \mathscr{L}_T^{\rightarrow}$, any two occurrences of a node labelled with φ, no matter in which semantic graph they occur, generate sub-graphs of their respective semantic graphs that are *structurally similar* between them. Of course, loops and systems of loops (possibly combined in various ways) complicate the picture.

- By the previous items, all the evaluations of nodes labelled with the same sentence, no matter in which semantic graph they occur, use sub-graphs that are structurally similar between them, thus yielding identical results.

The first task, then, is to make precise the notion of nodes reachable from a certain node (within a semantic graph) and of the resulting sub-graph.

Definition 5.3.1. (Reachable nodes and reachable sub-graphs)

Let $\varphi \in \mathscr{L}_T^{\rightarrow}$ and $\mathtt{v} \in \mathtt{N}_\varphi$. Define the set $\mathtt{R}_\varphi(\mathtt{v})$ of the tuples of ordered pairs of nodes *reachable from* $\mathtt{v}$ as follows:

$$\mathtt{R}_\varphi(\mathtt{v}) := \{\Big\langle \langle \mathtt{v},\mathtt{w}\rangle,\ldots,\langle \mathtt{v}',\mathtt{w}'\rangle \Big\rangle \in \bigcup_{n\in\omega}(\mathtt{N}_\varphi \times \mathtt{N}_\varphi)^n |$$
$$\text{there is a path from } \mathtt{v} \text{ to } \mathtt{w}' \text{ within } \mathtt{S}_\varphi\}.$$

Let $\varphi \in \mathscr{L}_T^{\rightarrow}$ and $\mathtt{v} \in \mathtt{N}_\varphi$. The *sub-graph of* $\langle \mathtt{N}_\varphi, \mathtt{S}_\varphi\rangle$ *reachable from* $\mathtt{v}$ is the graph $\langle \mathtt{N}^{\mathtt{v}}_\varphi, \mathtt{S}^{\mathtt{v}}_\varphi\rangle$ defined as follows:

$$\langle \mathtt{N}^{\mathtt{v}}_\varphi, \mathtt{S}^{\mathtt{v}}_\varphi\rangle := \langle \text{the nodes in } \mathtt{R}_\varphi(\mathtt{v}), \text{the edges in } \mathtt{R}_\varphi(\mathtt{v})\rangle.^{20}$$

▲

The second task is to devise a suitable notion of *structural similarity* for sub-graphs of semantic graphs, which takes into account all the factors used in the definition of e_φ, thus being usable to prove results on the evaluations of nodes occurring in structurally similar sub-graphs.

Definition 5.3.2. (*l*-isomorphism)
Let $\varphi,\psi \in \mathscr{L}_T^{\rightarrow}$, and let $\langle N^\dagger, S^\dagger\rangle \subseteq_g \langle N_\varphi, S_\varphi\rangle$ and $\langle N^{\dagger\dagger}, S^{\dagger\dagger}\rangle \subseteq_g \langle N_\psi, S_\psi\rangle$, where both $\langle N^\dagger, S^\dagger\rangle$ and $\langle N^{\dagger\dagger}, S^{\dagger\dagger}\rangle$ are labelled with $\mathscr{L}_T^{\rightarrow}$-sentences. Let's say that $\langle N^\dagger, S^\dagger\rangle$ and $\langle N^{\dagger\dagger}, S^{\dagger\dagger}\rangle$ are *l-isomorphic*, in symbols $\langle N^\dagger, S^\dagger\rangle \cong_l \langle N^{\dagger\dagger}, S^{\dagger\dagger}\rangle$, if all the following conditions are met:

1. For every simple leaf (simple point) $\mathtt{v}$ of $N^\dagger$ there is a simple leaf (simple point) $\mathtt{w}$ of $N^{\dagger\dagger}$ such that $\mathtt{v}$ and $\mathtt{w}$ have the same label, and dually for $N^{\dagger\dagger}$.

[20] I give this semi-formal definition for readability. Officially,

$$\mathtt{N}^{\mathtt{v}}_\varphi := \{\mathtt{w} \in \mathtt{N}_\varphi \,|\, \text{there is a } \langle \mathtt{v_i},\mathtt{v_j}\rangle \in \mathtt{S}_\varphi \text{ such that } \mathtt{w} = \mathtt{v_i} \text{ or } \mathtt{w} = \mathtt{v_j} \text{ and there is a path } \mathtt{P} \in \mathtt{R}_\varphi(\mathtt{v}) \text{ such that } \langle \mathtt{v_i},\mathtt{v_j}\rangle \in \mathtt{P}\}.$$
$$\mathtt{S}^{\mathtt{v}}_\varphi := \{\langle \mathtt{v_i},\mathtt{v_j}\rangle \in \mathtt{S}_\varphi \,|\, \text{ there is a path } \mathtt{P} \in \mathtt{R}_\varphi(\mathtt{v}) \text{ such that } \langle \mathtt{v_i},\mathtt{v_j}\rangle \in \mathtt{P}\}.$$

2. For each loop $\mathtt{L}_1$ within $\langle N^{\dagger}, S^{\dagger}\rangle$ there is a loop $\mathtt{L}_2$ within $\langle N^{\dagger\dagger}, S^{\dagger\dagger}\rangle$ s.t. the labels of nodes in $\mathtt{L}_1$ are identical to the labels of nodes in $\mathtt{L}_2$, and dually for loops within $\langle N^{\dagger\dagger}, S^{\dagger\dagger}\rangle$.

▲

Informally, two sub-graphs $\langle N^{\dagger}, S^{\dagger}\rangle$ and $\langle N^{\dagger\dagger}, S^{\dagger\dagger}\rangle$ are l-isomorphic when, besides satisfying condition 1 of the definition, every loop of $\langle N^{\dagger}, S^{\dagger}\rangle$ is mapped to a loop of $\langle N^{\dagger\dagger}, S^{\dagger\dagger}\rangle$ that features the same number of nodes with the same labels but that *can be oriented differently* (and *vice versa*). Since loops may appear in various combinations in semantic graphs (as discussed in Subsection 5.1.1), entire systems of loops (possibly having infinitely many loops) in l-isomorphic sub-graphs may present different rotations of the loops they are made of, thus making the graphical realizations of l-isomorphic sub-graphs look very different.

Remark 5.3.3.

l-isomorphism differs from the usual graph-theoretic isomorphism, i.e. the existence of a bijection preserving the adjacency of nodes in unlabelled graphs. To see the difference, consider λ and the sentence Tt', for $t' = \ulcorner Tt'' \urcorner$ and $t'' = \ulcorner Tt' \urcorner$; Tt' might be seen as a "double truth-teller". Now, $\langle N_{\lambda}, S_{\lambda}\rangle \ncong_l \langle N_{Tt'}, S_{Tt'}\rangle$ while $\langle N_{\lambda}, S_{\lambda}\rangle$ and $\langle N_{Tt'}, S_{Tt'}\rangle$ are isomorphic in the usual sense; such sense, then, is evidently too coarse-grained for our purposes of proving that the evaluations on nodes are uniform, as it does not take labels into account.

One could strengthen the usual notion of graph-theoretic isomorphism thus: two sub-graphs of semantic graphs are $+$*-isomorphic* if there is a bijection between them that preserves adjacency and identity of labels. This notion, however, it too fine-grained for our purposes. For example, if we want to give the same value to every sentence labelling nodes in $\mathtt{N}_{\mu}$ (see Figure 5.6) following the strategy sketched at the beginning of this subsection, we need to declare (also) that if φ

and ψ label nodes in $\mathtt{N}_{\mu}$, then $\langle \mathtt{N}_{\varphi}, \mathtt{S}_{\varphi} \rangle$ and $\langle \mathtt{N}_{\psi}, \mathtt{S}_{\psi} \rangle$ are relevantly similar in their structure to support evaluations that work in the same way and assign the same value to their nodes. l-isomorphism does the job, since for every φ and ψ as above $\langle \mathtt{N}_{\varphi}, \mathtt{S}_{\varphi} \rangle \cong_l \langle \mathtt{N}_{\psi}, \mathtt{S}_{\psi} \rangle$, but $+$-isomorphism does not, since there are sentences φ^+ and ψ^+ of the above kind such that $\langle \mathtt{N}_{\varphi^+}, \mathtt{S}_{\varphi^+} \rangle$ and $\langle \mathtt{N}_{\psi^+}, \mathtt{S}_{\psi^+} \rangle$ are not $+$-isomorphic (e.g. let φ^+ be μ and ψ^+ be $T \ulcorner T \ulcorner \mu \urcorner \urcorner$). ▲

The main structural fact on semantic graphs and their l-isomorphism properties, bridging Definition 5.3.1 and Definition 5.3.2, can now be stated.

Theorem 5.3.4. (Nodes with identical labels yield l-isomorphic reachable graphs)
Let $\varphi, \psi \in \mathscr{L}_T^{\rightarrow}$, $\mathtt{v} \in \mathtt{N}_{\varphi}$, $\mathtt{w} \in \mathtt{N}_{\psi}$. If $l_{\varphi}(\mathtt{v}) = l_{\psi}(\mathtt{w})$, then $\langle \mathtt{N}^{\mathtt{v}}_{\varphi}, \mathtt{S}^{\mathtt{v}}_{\varphi} \rangle \cong_l \langle \mathtt{N}^{\mathtt{w}}_{\psi}, \mathtt{S}^{\mathtt{w}}_{\psi} \rangle$.

Proof.
Let me start with two auxiliary results (whose proof is immediate).

Lemma 5.3.5.
Let $\varphi, \psi \in \mathscr{L}_T^{\rightarrow}$, $\mathtt{v} \in \mathtt{N}_{\varphi}$, $\mathtt{w} \in \mathtt{N}_{\psi}$. If $l_{\varphi}(\mathtt{v}) = l_{\psi}(\mathtt{w})$ and the clause (Loop) *does not apply to* v *nor to* w*, then* v *and* w *have the same number of immediate successors, with identical labels.*

Corollary 5.3.6.
Let $\varphi, \psi \in \mathscr{L}_T^{\rightarrow}$, $\mathtt{v} \in \mathtt{N}_{\varphi}$, $\mathtt{w} \in \mathtt{N}_{\psi}$. If $l_{\varphi}(\mathtt{v}) = l_{\psi}(\mathtt{w})$ and there is a $\mathtt{v}' \in \mathtt{Pred}_{\varphi}(\mathtt{v})$ s.t. (Loop) *applies to* v *and* $\mathtt{v}'$ *but there is no $\mathtt{w}' \in \mathtt{Pred}_{\psi}(\mathtt{w})$ s.t. $l(\mathtt{v}') = l(\mathtt{w}')$ and* (Loop) *applies to* w *and* $\mathtt{w}'$*, then* w *has an immediate successor* $\mathtt{w}''$ *s.t. $l(\mathtt{v}') = l(\mathtt{w}'')$.*

I will only prove the claims concerning $\langle \mathtt{N}^{\mathtt{v}}_{\varphi}, \mathtt{S}^{\mathtt{v}}_{\varphi} \rangle$, as the other direction is dual.

(a) Let v be a simple leaf: $l_{\varphi}(\mathtt{v})$ is an atomic arithmetical sentence or a sentence of the form Tt where t does not denote the code of a $\mathscr{L}_T^{\rightarrow}$-sentence. $l_{\varphi}(\mathtt{v}) = l_{\psi}(\mathtt{w})$, so $l_{\psi}(\mathtt{w})$ is an atomic arithmetical sentence or with Tt for a term t that does not denote the code of a $\mathscr{L}_T^{\rightarrow}$-sentence. By Definition 5.1.2, w is a simple leaf.

(b) Suppose that $\mathtt{v}$ is a simple point and, by contradiction, that $\mathtt{w}$ is not a simple point. As we have just seen, $\mathtt{w}$ is not a simple leaf, so it belongs to at least one loop $\mathtt{L}$. By definition, $\mathtt{L}$ has the following form:[21]

$$\mathtt{L} = \Big\langle \langle \mathtt{w_1}, \mathtt{w_2} \rangle, \ldots \langle \mathtt{w_i}, \mathtt{w} \rangle, \langle \mathtt{w}, \mathtt{w_{i+1}} \rangle, \ldots, \langle \mathtt{w_j}, \mathtt{w_{j+1}} \rangle, \langle \mathtt{w_{j+1}}, \mathtt{w_1} \rangle \Big\rangle,$$

where the clause $(\mathtt{Loop})$ applies to $\mathtt{w_{j+1}}$ (and possibly other nodes as well). Since $l_\varphi(\mathtt{v}) = l_\psi(\mathtt{w})$, by Lemma 5.3.5 and Corollary 5.3.6, $\mathtt{v}$ and $\mathtt{w}$ have the same number of successor nodes, labelled with the same sentences. $\mathtt{v}$ and $\mathtt{w}$ can have many successor nodes, possibly infinitely many, but in particular there will be a node $\mathtt{v_1}$ s.t. $\langle \mathtt{v}, \mathtt{v_1} \rangle \in \mathtt{S}^{\mathtt{v}}_\varphi$ and $l_\psi(\mathtt{w_{i+1}}) = l_\varphi(\mathtt{v_2})$. By the assumption that $\mathtt{v}$ is a simple point, we see that $\mathtt{v_1}$ cannot loop back to $\mathtt{v}$ nor to any of its predecessors, otherwise $\mathtt{v}$ would not be a simple point, given that $\mathtt{v} \in \mathtt{Pred}_\varphi(\mathtt{v_1})$. Moreover, it is not the case that $(\mathtt{Loop})$ applies to $\mathtt{v_1}$ with itself *and* it is the only clause that applies to it (namely that $\mathtt{v_1}$ *only* loops back to itself), because otherwise $\mathtt{w_{i+1}}$ would do that as well (by the above lemma and corollary), and it does not. So, $\mathtt{v_1}$ is also a simple point. Now we apply to $\mathtt{v_1}$ and $\mathtt{w_{i+1}}$ the same reasoning that we applied to $\mathtt{v}$ and $\mathtt{w}$: using Lemma 5.3.5 and Corollary 5.3.6 on $\mathtt{v_1}$ and $\mathtt{w_{i+1}}$ we see that, between possibly other ones, $\mathtt{v_1}$ has at least one successor that is labelled as the successor node of $\mathtt{w_{i+1}}$ that belongs to $\mathtt{L}$. For the same reason as above, such successor of $\mathtt{v_1}$ is a simple point. Proceeding in this fashion, by repeated applications of Lemma 5.3.5 and Corollary 5.3.6, we generate a path $\mathtt{P}$ within $\mathtt{S}^{\mathtt{v}}_\varphi$ s.t., as set of ordered pairs of sentences (labels), matches exactly the loop $\mathtt{L}$.[22] So, $\mathtt{P}$ will have the form:

$$\mathtt{P} = \Big\langle \langle \mathtt{v}, \mathtt{v_1} \rangle, \langle \mathtt{v_1}, \mathtt{v_2} \rangle, \ldots, \langle \mathtt{v_k}, \mathtt{v_{k+1}} \rangle \Big\rangle,$$

[21] In this proof, I suppose for the sake of readability and without loss of generality that the nodes in the mentioned loops and paths are indexed progressively.

[22] Corollary 5.3.6 is used in generating the successor of $\mathtt{v}$ (within $\mathtt{P}$) labelled as $\mathtt{w_{j+1}}$

where $l_\varphi(\mathtt{v}) = l_\psi(\mathtt{w}), \ldots, l_\varphi(\mathtt{v_{k+1}}) = l_\psi(\mathtt{w_i})$ (the dots stand for the nodes in $\mathtt{L}$) and $j = k$. However, by our assumption that $\mathtt{v}$ is a simple point, the clause $(\mathtt{Loop})$ never applies to a node in $\mathtt{P}$ and $\mathtt{v}$ or one of its predecessors. But, by Corollary 5.3.6 and Definition 5.1.2, the clause $(\mathtt{Loop})$ applies at least once to $\mathtt{v_{k+1}}$ that loops back to $\mathtt{v}$. Contradiction. So, if $\mathtt{v}$ is a simple point, also $\mathtt{w}$ is a simple point. They have the same label by assumption and, as shown, they have the same number of immediate successors, labelled with the same sentences, by Lemma 5.3.5.

(c) Let $\mathtt{v}$ belong to at least one loop. Suppose by contradiction that there is at least one loop $\mathtt{L}^*$ s.t. $\mathtt{v}$ belongs to $\mathtt{L}^*$ and there is no loop $\mathtt{L}$ within $\langle \mathtt{N}^{\mathtt{w}}_\psi, \mathtt{S}^{\mathtt{w}}_\psi \rangle$ having exactly as many nodes as $\mathtt{L}^*$ and labelled in the same way. We will derive a contradiction from this assumption, by showing that there is a path within $\langle \mathtt{N}^{\mathtt{w}}_\psi, \mathtt{S}^{\mathtt{w}}_\psi \rangle$ that has the same number of nodes of $\mathtt{L}^*$, labelled in the same way – equivalently, that is l-isomorphic to $\mathtt{L}^*$. For notational convenience, from now on, if $\mathtt{Q}$ is a set of edges or a tuple of edges within a graph $\langle \mathtt{N}_\chi, \mathtt{S}_\chi \rangle$, I will use $l_\chi(\mathtt{Q})$ to denote the set of $\mathscr{L}_T^{\rightarrow}$-sentences labelling nodes in $\mathtt{Q}$.

$\mathtt{L}^*$ is a finite tuple of ordered pairs of labelled nodes of the following form:

$$\mathtt{L}^* = \Big\langle \langle \mathtt{v_1}, \mathtt{v_2} \rangle, \ldots \langle \mathtt{v_i}, \mathtt{v} \rangle, \langle \mathtt{v}, \mathtt{v_{i+1}} \rangle, \ldots, \langle \mathtt{v_k}, \mathtt{v_{k+1}} \rangle, \langle \mathtt{v_{k+1}}, \mathtt{v_1} \rangle \Big\rangle,$$

where the clause $(\mathtt{Loop})$ applies to $\mathtt{v_{k+1}}$ that loops back to $\mathtt{v_1}$. By our supposition, there is no loop $\mathtt{L}$ in $\langle \mathtt{N}^{\mathtt{w}}_\psi, \mathtt{S}^{\mathtt{w}}_\psi \rangle$ that has the same number of nodes of $\mathtt{L}^*$ and such that $l_\psi(\mathtt{L}) = l_\varphi(\mathtt{L}^*)$. By (a) and (b) above, $\mathtt{w}$ belongs to at least one loop. Moreover, Lemma 5.3.5 and Corollary 5.3.6 imply that $\mathtt{w}$ has (within $\langle \mathtt{N}^{\mathtt{w}}_\psi, \mathtt{S}^{\mathtt{w}}_\psi \rangle$) the same number of immediate successors that $\mathtt{v}$ has (within $\langle \mathtt{N}^{\mathtt{v}}_\varphi, \mathtt{S}^{\mathtt{v}}_\varphi \rangle$), with the same labels, and this fact applies iteratively to all their successors as well. So, if $\mathtt{w}$ is not contained in any loop that is l-isomorphic to $\mathtt{L}^*$, then one of the following must be the case:

(c.1) the clause $(\mathtt{Loop})$ applies to $\mathtt{w}$ or to one of its successors (within $\langle \mathtt{N}^{\mathtt{w}}_\psi, \mathtt{S}^{\mathtt{w}}_\psi \rangle$)

that is labelled as a node in $\mathtt{L}^*$ but not to such node in $\mathtt{L}^*$;

(c.2) the clause $(\mathtt{Loop})$ applies to $\mathtt{v}$ or to one of its successors (within $\langle \mathtt{N}^{\mathtt{v}}_{\varphi}, \mathtt{S}^{\mathtt{v}}_{\varphi}\rangle$) that are in $\mathtt{L}^*$ but not to any node amongst $\mathtt{w}$ and its successors (within $\langle \mathtt{N}^{\mathtt{w}}_{\psi}, \mathtt{S}^{\mathtt{w}}_{\psi}\rangle$) having the same label.

Suppose that (c.1) is the case – otherwise, the proof proceeds in a dual way. (c.1) is in fact a disjunctive condition, that may be presented as follows:

(c.1.1) the clause $(\mathtt{Loop})$ applies to $\mathtt{w}$ but not to any node in $\mathtt{L}^*$ labelled as $\mathtt{w}$;

(c.1.2) the clause $(\mathtt{Loop})$ applies to one of the successors of $\mathtt{w}$ that is labelled as a node in $\mathtt{L}^*$ but not to such node in $\mathtt{L}^*$.

I will suppose that (c.1.2) is the case. If (c.1.1) holds, instead, the following reasoning still applies (it is slightly simpler, in fact), only with notational changes.

Let $\mathtt{w}_{\mathtt{p}+1}$ be the first successor of $\mathtt{w}$ s.t. $(\mathtt{Loop})$ applies to $\mathtt{w}_{\mathtt{p}+1}$ but not to the node labelled as $\mathtt{w}_{\mathtt{p}+1}$ in $\mathtt{L}^*$. By the *first* successor of $\mathtt{w}$ that respects the above conditions, I mean the *closest successor* of $\mathtt{w}$ that does.[23] If there are many minimally close successors respecting the above condition, pick any of them – the following argument does not depend on uniqueness. So, there is a path $\mathtt{P}$ s.t.

$$\Big\langle \langle \mathtt{w}, \mathtt{w}_{\mathtt{n}+1}\rangle, \dots, \langle \mathtt{w}_{\mathtt{p}}, \mathtt{w}_{\mathtt{p}+1}\rangle, \langle \mathtt{w}_{\mathtt{p}+1}, \mathtt{w}_{\mathtt{m}}\rangle \Big\rangle = \mathtt{P} \in \mathtt{R}_{\psi}(\mathtt{w}),$$

where, by Lemma 5.3.5 and Corollary 5.3.6, $l_{\psi}(\mathtt{w}_{\mathtt{n}+1}) = l_{\varphi}(\mathtt{v}_{\mathtt{i}+1})$, ... , $l_{\psi}(\mathtt{w}_{\mathtt{p}}) = l_{\varphi}(\mathtt{v}_{\mathtt{i}+(\mathtt{p}-\mathtt{n})})$, and $l_{\psi}(\mathtt{w}_{\mathtt{p}+1}) = l_{\varphi}(\mathtt{v}_{\mathtt{i}+(\mathtt{p}-\mathtt{n})+1})$. The dots range over nodes of $\mathtt{P}$ that are labelled as nodes in $\mathtt{L}^*$ (they exist by the above lemma and corollary) and the clause $(\mathtt{Loop})$ applies to $\mathtt{w}_{\mathtt{p}+1}$ and some $\mathtt{w}_{\mathtt{m}}$ labelled as $\mathtt{v}_{\mathtt{i}+(\mathtt{p}-\mathtt{n})+2}$ (by our assumption), but it does not apply to $\mathtt{v}_{\mathtt{i}+(\mathtt{p}-\mathtt{n})+1}$ and $\mathtt{v}_{\mathtt{i}+(\mathtt{p}-\mathtt{n})+2}$ (by our assumption). It

[23]By *closest successor*, I mean the successor of $\mathtt{w}$ that respects the above condition and that is separated from $\mathtt{w}$ by the smallest number of nodes – the successor $\mathtt{w}^*$ of $\mathtt{w}$ that respects the above condition and such that the shortest path from $\mathtt{w}$ to $\mathtt{w}^*$ is shorter than any other path from $\mathtt{w}$ to a successor of it that respects the above condition.

follows that $\mathtt{v}_{\mathtt{i}+(\mathtt{p}-\mathtt{n})+1}$ is between $\mathtt{v}_{\mathtt{i}+1}$ and $\mathtt{v}_{\mathtt{k}}$ in $\mathtt{L}^*$, so there is an index $\mathtt{l}$ s.t. $\mathtt{v}_{\mathtt{i}+(\mathtt{p}-\mathtt{n})+1}$ is $\mathtt{v}_{\mathtt{l}}$. So, $\mathtt{L}^*$ looks as follows:

$$\mathtt{L}^* = \Big\langle \langle \mathtt{v}_1, \mathtt{v}_2 \rangle, \dots \langle \mathtt{v}_\mathtt{i}, \mathtt{v} \rangle, \langle \mathtt{v}, \mathtt{v}_{\mathtt{i}+1} \rangle, \dots, \langle \mathtt{v}_\mathtt{l}, \mathtt{v}_{\mathtt{l}+1} \rangle, \dots, \langle \mathtt{v}_\mathtt{k}, \mathtt{v}_{\mathtt{k}+1} \rangle, \langle \mathtt{v}_{\mathtt{k}+1}, \mathtt{v}_1 \rangle \Big\rangle.$$

Since $(\mathtt{Loop})$ applies to $\mathtt{w}_{\mathtt{p}+1}$, there is a path $\mathtt{P}_1$ s.t.

$$\mathtt{P} \subsetneq_p \Big\langle \langle \mathtt{w}_\mathtt{m}, \mathtt{w}_{\mathtt{m}+1} \rangle, \dots, \langle \mathtt{w}, \mathtt{w}_{\mathtt{n}+1} \rangle, \dots, \langle \mathtt{w}_\mathtt{p}, \mathtt{w}_{\mathtt{p}+1} \rangle, \langle \mathtt{w}_{\mathtt{p}+1}, \mathtt{w}_\mathtt{m} \rangle \Big\rangle = \mathtt{P}_1 \in \mathtt{R}_\psi(\mathtt{w}),$$

where $l_\psi(\mathtt{w}_\mathtt{m}) = l_\varphi(\mathtt{v}_{\mathtt{l}+1})$, $l(\mathtt{w}_{\mathtt{m}+1}) = l(\mathtt{v}_{\mathtt{l}+2})$. Our assumption dictates that there is no path within $\langle \mathtt{N}^\mathtt{w}_\psi, \mathtt{S}^\mathtt{w}_\psi \rangle$ that is l-isomorphic to $\mathtt{L}^*$. So, any path within $\langle \mathtt{N}^\mathtt{w}_\psi, \mathtt{S}^\mathtt{w}_\psi \rangle$ that contains nodes labelled as $l(\langle \mathtt{w}_\mathtt{m}, \mathtt{w}_{\mathtt{m}+1} \rangle \cup \mathtt{P})$ is not l-isomorphic to $\mathtt{L}^*$.[24] There are at most countably many such paths: call them $\mathtt{P}_{1.1}, \mathtt{P}_{1.2}, \dots$ ($\mathtt{P}_1$ is one of them). Considering all the (at most countably infinitely many) possible cases, we have:

$$\begin{aligned}
\Big\langle \langle \mathtt{w}_\mathtt{m}, \mathtt{w}_{\mathtt{m}+1} \rangle, \sim_1, \langle \mathtt{w}, \mathtt{w}_{\mathtt{n}+1} \rangle, \dots, \langle \mathtt{w}_{\mathtt{p}+1}, \mathtt{w}_\mathtt{m} \rangle \Big\rangle &= \mathtt{P}_{1.1} \in \mathtt{R}_\psi(\mathtt{w}) \\
&\vdots \\
\Big\langle \langle \mathtt{w}_\mathtt{m}, \mathtt{w}_{\mathtt{m}+1} \rangle, \sim_j, \langle \mathtt{w}^\mathtt{j}, \mathtt{w}^\mathtt{j}_{\mathtt{n}+1} \rangle, \dots, \langle \mathtt{w}^\mathtt{j}_{\mathtt{p}+1}, \mathtt{w}_\mathtt{m} \rangle \Big\rangle &= \mathtt{P}_{1.\mathtt{j}} \in \mathtt{R}_\psi(\mathtt{w}) \\
&\vdots
\end{aligned}$$

where $\sim_j$ indicates some (possibly empty) path within $\langle \mathtt{N}^\mathtt{w}_\psi, \mathtt{S}^\mathtt{w}_\psi \rangle$ that is *different* in labels from $l_\varphi(\mathtt{L}^*) \setminus [l_\psi(\langle \mathtt{w}_\mathtt{m}, \mathtt{w}_{\mathtt{m}+1} \rangle) \cup \bigcap_i l_\psi(\mathtt{P}_{1.\mathtt{i}})]$; furthermore, for every $i \in \omega$, we have that $l_\psi(\mathtt{w}) = l_\psi(\mathtt{w}^\mathtt{i})$, $l_\psi(\mathtt{w}_{\mathtt{n}+1}) = l_\psi(\mathtt{w}^\mathtt{i}_{\mathtt{n}+1})$, ..., $l_\psi(\mathtt{w}_{\mathtt{p}+1}) = l_\psi(\mathtt{w}^\mathtt{i}_{\mathtt{p}+1})$.

Since again Lemma 5.3.5 and Corollary 5.3.6 imply that $\mathtt{w}_{\mathtt{m}+1}$ has as many successors as $\mathtt{v}_{\mathtt{l}+1}$, and labelled in the same way (and the same holds for the successors of $\mathtt{v}_{\mathtt{l}+1}$), there is a node in $\mathtt{N}^{\mathtt{w}_{\mathtt{m}+1}}_\psi$ labelled as some member of $\mathtt{L}^*$ s.t. the

[24]In other terms, any path that agrees with $\mathtt{P}_1$ on $l(\langle \mathtt{w}_\mathtt{m}, \mathtt{w}_{\mathtt{m}+1} \rangle \cup \mathtt{P})$ disagrees with $l(\mathtt{L}^*)$ on the remaining pairs of $l(\mathtt{L}^*)$, i.e. on $l(\mathtt{L}^*) \setminus l(\langle \mathtt{w}_\mathtt{m}, \mathtt{w}_{\mathtt{m}+1} \rangle \cup \mathtt{P})$.

clause (Loop) applies to it and not to the corresponding member of $\mathtt{L}^*$, or *vice versa* (again because otherwise we would have a loop within $\langle \mathtt{N}^{\mathtt{w}}_{\psi}, \mathtt{S}^{\mathtt{w}}_{\psi} \rangle$ that would be l-isomorphic to $\mathtt{L}^*$). In terms of the above list, this means that some node in some of the $\mathtt{P}_{1.n}$ in the disagreeing part of the path $\sim_n$ loops back to a predecessor of $\mathtt{w}_{\mathtt{m}}$, and thus such node is a predecessor of all nodes that are successors of $\mathtt{w}_{\mathtt{m}}$.[25]

Take any such path, call it $\mathtt{P}_2$. We have that:

$$\Big\langle \langle \mathtt{w}_{\mathtt{o}}, \mathtt{w}_{\mathtt{o}+1} \rangle, \ldots, \langle \mathtt{w}_{\mathtt{q}}, \mathtt{w}_{\mathtt{q}+1} \rangle, \langle \mathtt{w}_{\mathtt{q}+1}, \mathtt{w}_{\mathtt{o}} \rangle \Big\rangle = \mathtt{P}_2 \in \mathtt{R}_{\psi}(\mathtt{w}),$$

for $l_{\psi}(\mathtt{w}_{\mathtt{o}}) = l_{\varphi}(\text{some node in } \mathtt{L}^*)$, $l_{\psi}(\mathtt{w}_{\mathtt{o}+1}) = l_{\varphi}(\text{its successor in } \mathtt{L}^*)$; moreover, (Loop) applies to $\mathtt{w}_{\mathtt{q}+1}$ and $\mathtt{w}_{\mathtt{o}}$. Clearly, for all $n \in \omega$, we have that $l_{\psi}(\mathtt{w}_{\mathtt{q}+1}) \neq l_{\psi}(\mathtt{w}^{\mathtt{n}}_{\mathtt{p}+1})$ (by construction and because otherwise no path such as $\mathtt{P}_{1.\mathtt{n}}$ could exist). By construction, $\mathtt{w}_{\mathtt{o}}$ and $\mathtt{w}_{\mathtt{o}+1}$ are predecessors of any node in $\bigcup_i \mathtt{P}_{1.\mathtt{i}}$. Using our assumption again, we reason as we did after finding the path $\mathtt{P}_1$, deriving the existence of (possibly countably infinitely many) paths $\mathtt{P}_{2.1}, \mathtt{P}_{2.2}, \ldots$ s.t.:

$$\begin{aligned}
\Big\langle \langle \mathtt{w}_{\mathtt{o}}, \mathtt{w}_{\mathtt{o}+1} \rangle, \sim_1, \langle \mathtt{w}_{\mathtt{q}}, \mathtt{w}_{\mathtt{q}+1} \rangle, \ldots, \langle \mathtt{w}_{\mathtt{q}+1}, \mathtt{w}_{\mathtt{o}} \rangle \Big\rangle &= \mathtt{P}_{2.1} \in \mathtt{R}_{\psi}(\mathtt{w}) \\
&\vdots \\
\Big\langle \langle \mathtt{w}_{\mathtt{o}}, \mathtt{w}_{\mathtt{o}+1} \rangle, \sim_j, \langle \mathtt{w}^{\mathtt{j}}_{\mathtt{q}}, \mathtt{w}^{\mathtt{j}}_{\mathtt{q}+1} \rangle, \ldots, \langle \mathtt{w}^{\mathtt{j}}_{\mathtt{q}+1}, \mathtt{w}^{\mathtt{j}}_{\mathtt{o}} \rangle \Big\rangle &= \mathtt{P}_{2.\mathtt{j}} \in \mathtt{R}_{\psi}(\mathtt{w}) \\
&\vdots
\end{aligned}$$

$\sim_j$ indicates a path *different* in labels from $l_{\varphi}(\mathtt{L}^*) \setminus l_{\psi}[(\langle \mathtt{w}_{\mathtt{o}}, \mathtt{w}_{\mathtt{o}+1} \rangle) \cup \bigcap_i l_{\psi}(\mathtt{P}_{2.\mathtt{i}})]$. Then we take an arbitrary path $\mathtt{P}_3$, as we did for $\mathtt{P}_2$, i.e. s.t. the clause (Loop) applies to two nodes in it labelled as in $\mathtt{L}^*$, but s.t. (by assumption) the resulting path is not l-isomorphic to $\mathtt{L}^*$. By construction, the node labelled as a node in $\mathtt{L}^*$ that loops back to a node labelled as a node in $\mathtt{L}^*$ is s.t. no node with the same

[25] As above, such node exists Lemma 5.3.5 and Corollary 5.3.6, whereas the conclusion about the clause (Loop) is forced by our assumption.

label loops back to a node labelled as a node in $\mathtt{L}^*$ in $\mathtt{P}_1$ or $\mathtt{P}_2$ (just as above we had that $l(\mathtt{w}_{\mathtt{q}+1}) \neq l(\mathtt{w}^{\mathtt{n}}_{\mathtt{p}+1})$ for all $n \in \omega$). Moreover, the node in $\mathtt{P}_3$ to which such node loops back is a predecessor of every node in $(\bigcup_i \mathtt{P}_{1.\mathtt{i}} \cup \bigcup_i \mathtt{P}_{2.\mathtt{i}})$ (just as we had that $\mathtt{w}_{\mathtt{o}}$ and $\mathtt{w}_{\mathtt{o}+1}$ are predecessors of every node in $\bigcup_i \mathtt{P}_{1.\mathtt{i}}$ above).

Proceeding in this fashion, we derive the existence of more paths $\mathtt{P}_4, \mathtt{P}_5, \ldots$ labelled as subsets of $l(\mathtt{L}^*)$: by our assumption, no $\mathtt{P}_{\mathtt{t}+1}$ in $\mathtt{P}_1$, $\mathtt{P}_2$, $\mathtt{P}_3$, $\mathtt{P}_4$, $\mathtt{P}_5$, $\ldots$ is l-isomorphic to $\mathtt{L}^*$; our assumption, Lemma 5.3.5, and Corollary 5.3.6 force us to conclude that each $\mathtt{P}_{\mathtt{t}+1}$ is labelled as a proper subset of $\mathtt{L}^*$ but is not l-isomorphic to it. The existence of such nodes is guaranteed by Lemma 5.3.5 and Corollary 5.3.6, and the "asymmetric" looping back is required by our assumption that no such path can be l-isomorphic to $\mathtt{L}^*$: if no node in such a path looped back asymmetrically, a contradiction would be immediate.

The key facts are these: (1) for each path $\mathtt{P}_{\mathtt{t}+1}$ between $\mathtt{P}_1, \mathtt{P}_2, \mathtt{P}_3, \mathtt{P}_4, \ldots$ the clause $(\mathtt{Loop})$ applies to a node in it labelled as a node in $\mathtt{L}^*$ and no node with the same label in any of $\mathtt{P}_1, \mathtt{P}_2, \mathtt{P}_3, \mathtt{P}_4, \ldots, \mathtt{P}_{\mathtt{t}}$ is s.t. $(\mathtt{Loop})$ applies to it and to a node labelled as a node in $\mathtt{L}^*$ (by construction and to avoid contradiction, as sen above); (2) for each path $\mathtt{P}_{\mathtt{t}+1}$ between $\mathtt{P}_1, \mathtt{P}_2, \mathtt{P}_3, \mathtt{P}_4, \ldots$ a node labelled as an element of $\mathtt{L}^*$ loops back to a node that is also labelled as an element of $\mathtt{L}^*$: the latter is a predecessor of every node in $\bigcup_{m \leq t}(\bigcup_i \mathtt{P}_{\mathtt{m}.\mathtt{i}})$ (by the definition of predecessor and the above construction (which in turn is forced by Lemma 5.3.5, Corollary 5.3.6 and our assumption) as seen above).

But $\mathtt{L}^*$ has only finitely many nodes, so this process cannot go on forever. Deriving the existence of finitely many $\mathtt{P}_1, \mathtt{P}_2, \mathtt{P}_3, \mathtt{P}_4, \ldots$, we should find a $\mathtt{P}_{\mathtt{n}+1}$

$$\Big\langle \langle \mathtt{w}_{\mathtt{r}}, \mathtt{w}_{\mathtt{r}+1} \rangle, \ldots, \langle \mathtt{w}_{\mathtt{s}}, \mathtt{w}_{\mathtt{s}+1} \rangle, \langle \mathtt{w}_{\mathtt{s}+1}, \mathtt{w}_{\mathtt{r}} \rangle \Big\rangle = \mathtt{P}_{\mathtt{n}+1}$$

s.t. $l_{\psi}(\mathtt{w}_{\mathtt{s}+1})$ is identical to the label of some node between $\mathtt{v}_{\mathtt{i}+2}$ and $\mathtt{v}_{\mathtt{l}-1}$ (by

Lemma 5.3.5 and Corollary 5.3.6) and the clause $(\mathtt{Loop})$ applies to $\mathtt{w_{s+1}}$ and $\mathtt{w_r}$ (by our supposition and key fact (1)). But this cannot be! For, if there were a node labelled as $\mathtt{w_r}$ with which $\mathtt{w_{s+1}}$ is in a loop, then there would have already been some node between $\mathtt{w_{n+1}}$ and $\mathtt{w_p}$ (included) within $\mathtt{P_1}$ in a loop with it. This is because $\mathtt{w_r}$ is a predecessor of every node between $\mathtt{w_{n+1}}$ and $\mathtt{w_p}$ (included), by key fact (2). So, no path such as $\mathtt{P_1}$ could have existed, because the nodes between $\mathtt{w_{n+1}}$ and $\mathtt{w_p}$ are labelled as the nodes between $\mathtt{v_{i+2}}$ and $\mathtt{v_{l-1}}$ (included), and $\mathtt{w_{p+1}}$ would not be the first successor of $\mathtt{w}$ (i.e. the/a successor at minimal distance from $\mathtt{w}$) in $\mathtt{N}^{\mathtt{w}}_{\psi}$ labelled as a member of a pair in $\mathtt{L}^*$ such that the clause $(\mathtt{Loop})$ applies to it but not to the node with the same label in $\mathtt{L}^*$.[26] Then, we derive the existence of

$$\Big\langle \langle \mathtt{w_s}, \mathtt{w_{s+1}} \rangle, \langle \mathtt{w_{s+1}}, \mathtt{w_{s+2}} \rangle \Big\rangle = \widehat{\mathtt{P}}_1 \in \mathtt{R}_{\psi}(\mathtt{w}),$$

where $l_{\psi}(\mathtt{w_{s+1}}) = l_{\varphi}(\mathtt{v_{i+2}})$ (by Lemma 5.3.5 and Corollary 5.3.6) and $(\mathtt{Loop})$ does not apply to $\mathtt{w_{s+1}}$ and any other node labelled as in $\mathtt{L}^*$.

We reiterate this reasoning $k+1$ times, noting that the clause $(\mathtt{Loop})$ never applies to $\mathtt{w_{s+m}}$ (for $m \leq k+1$) and to some node in $\mathtt{N}^{\mathtt{w}}_{\psi}$ labelled as a node in $\mathtt{L}^*$, otherwise we would have a contradiction with the existence of the paths $\mathtt{P_1}, \mathtt{P_2}, \mathtt{P_3}, \mathtt{P_4}, \dots$ derived so far. This reiteration yields the existence of the following paths:

$$\begin{aligned} \Big\langle \langle \mathtt{w_s}, \mathtt{w_{s+1}} \rangle, \langle \mathtt{w_{s+1}}, \mathtt{w_{s+2}} \rangle, \langle \mathtt{w_{s+2}}, \mathtt{w_{s+3}} \rangle \Big\rangle &= \widehat{\mathtt{P}}_2 \in \mathtt{R}_{\psi}(\mathtt{w}) \\ &\vdots \\ \Big\langle \langle \mathtt{w_s}, \mathtt{w_{s+1}} \rangle, \langle \mathtt{w_{s+1}}, \mathtt{w_{s+2}} \rangle, \dots, \langle \mathtt{w_{s+k}}, \mathtt{w_{s+k+1}} \rangle, \langle \mathtt{w_{s+k+1}}, \mathtt{w_s} \rangle \Big\rangle &= \widehat{\mathtt{P}}_k \in \mathtt{R}_{\psi}(\mathtt{w}) \end{aligned}$$

By construction, Lemma 5.3.5, and Corollary 5.3.6, $\widehat{\mathtt{P}}_k$ has as many nodes as $\mathtt{L}^*$ and their labels are identical to those of the nodes in $\mathtt{L}^*$. Contradiction. □

[26] If case (c.1.1) is given, one should simply erase the latter sentence concerning $\mathtt{w_{p+1}}$.

Theorem 5.3.4 allows us to show that the evaluation functions of Definition 5.2.3 are uniform in the sense described above, captured by the following theorem.

Theorem 5.3.7. (Uniformity of the parametrized evaluations of nodes)
Let $\varphi, \psi \in \mathscr{L}_T^{\rightarrow}$, $\mathtt{v} \in \mathtt{N}_\varphi$, and $\mathtt{w} \in \mathtt{N}_\psi$ be s.t. $l_\varphi(\mathtt{v}) = l_\psi(\mathtt{w})$. Then, for $vl \in Ł \cup V$:

there is an ordinal α s.t. $e_\varphi(\alpha, \mathtt{v}) = vl$ iff there is an ordinal β s.t. $e_\psi(\beta, \mathtt{w}) = vl$.

Proof sketch.
Let φ, ψ, $\mathtt{v}$, and $\mathtt{w}$ be as in the theorem. First, let's prove a useful lemma.

Lemma 5.3.8.
Let $\varphi, \psi \in \mathscr{L}_T^{\rightarrow}$, $\mathtt{v} \in \mathtt{N}_\varphi$, and $\mathtt{w} \in \mathtt{N}_\psi$ be s.t. $l_\varphi(\mathtt{v}) = l_\psi(\mathtt{w})$. For every simple leaf or simple point $\mathtt{v_i} \in \mathtt{N}^{\mathtt{v}}_\varphi$ there is a simple leaf or simple point $\mathtt{w_j} \in \mathtt{N}^{\mathtt{w}}_\psi$ s.t.:

- $l_\varphi(\mathtt{v_i}) = l_\psi(\mathtt{w_j})$,
- $l_\varphi(\textit{the immediate predecessor of } \mathtt{v_i}) = l_\psi(\textit{the immediate predecessor of } \mathtt{w_j})$.

Proof of the lemma.
For $\mathtt{v_i}$ as in the lemma, the immediate predecessor of $\mathtt{v_i}$, call it $\mathtt{v_i^\star}$, is in $\mathtt{N}^{\mathtt{v}}_\varphi$. Since by Theorem 5.3.4 $\langle \mathtt{N}^{\mathtt{v}}_\varphi, \mathtt{S}^{\mathtt{v}}_\varphi \rangle \cong_l \langle \mathtt{N}^{\mathtt{w}}_\psi, \mathtt{S}^{\mathtt{w}}_\psi \rangle$, there is a $\mathtt{w}^\star \in \mathtt{N}^{\mathtt{w}}_\psi$ s.t. $l_\varphi(\mathtt{v_i^\star}) = l_\psi(\mathtt{w}^\star)$. Moreover, by Lemma 5.3.5 and Corollary 5.3.6, $\mathtt{w}^\star$ has within $\langle \mathtt{N}^{\mathtt{w}}_\psi, \mathtt{S}^{\mathtt{w}}_\psi \rangle$ the same number of successor nodes to which the clause $(\mathtt{Loop})$ does not apply that $\mathtt{v_i}^\star$ has within $\langle \mathtt{N}^{\mathtt{v}}_\varphi, \mathtt{S}^{\mathtt{v}}_\varphi \rangle$, and labelled in the same way. From this, the existence of a node $\mathtt{w_j}$ as in the statement of the lemma follows immediately – in fact, if $\mathtt{w_j}$ is not in a loop, its being a simple point or a simple leaf depends only on its label. $\square$

Proof sketch of the theorem (continued).
I will only do the left-to-right direction. Let there be an α s.t. $e_\varphi(\alpha, \mathtt{v}) = vl$. By Definition 5.2.3, the only nodes of $\mathtt{N}_\varphi$ used in establishing $e_\varphi(\alpha, \mathtt{v}) = vl$ are

those in $\mathtt{N}^{\mathtt{v}}_{\varphi}$. Since $l_{\varphi}(\mathtt{v}) = l_{\psi}(\mathtt{w})$, by Theorem 5.3.4 $\langle \mathtt{N}^{\mathtt{v}}_{\varphi}, \mathtt{S}^{\mathtt{v}}_{\varphi} \rangle \cong_{l} \langle \mathtt{N}^{\mathtt{w}}_{\psi}, \mathtt{S}^{\mathtt{w}}_{\psi} \rangle$. As a consequence:

(i) Simple leaves of $\langle \mathtt{N}^{\mathtt{v}}_{\varphi}, \mathtt{S}^{\mathtt{v}}_{\varphi} \rangle$ are mapped to simple leaves of $\langle \mathtt{N}^{\mathtt{w}}_{\psi}, \mathtt{S}^{\mathtt{w}}_{\psi} \rangle$ with identical labels, and *vice versa*.

(ii) Simple points of $\langle \mathtt{N}^{\mathtt{v}}_{\varphi}, \mathtt{S}^{\mathtt{v}}_{\varphi} \rangle$ are mapped to simple points of $\langle \mathtt{N}^{\mathtt{w}}_{\psi}, \mathtt{S}^{\mathtt{w}}_{\psi} \rangle$ with identical labels, and by Lemma 5.3.8 every path made of simple points within $\langle \mathtt{N}^{\mathtt{v}}_{\varphi}, \mathtt{S}^{\mathtt{v}}_{\varphi} \rangle$ (with the only possible exception of starting with a looping point or ending in a simple leaf) is reconstructed, identical in order and labels, within $\langle \mathtt{N}^{\mathtt{w}}_{\psi}, \mathtt{S}^{\mathtt{w}}_{\psi} \rangle$, and *vice versa*.

(iii) Every loop within $\langle \mathtt{N}^{\mathtt{v}}_{\varphi}, \mathtt{S}^{\mathtt{v}}_{\varphi} \rangle$ is mapped to a loop of $\langle \mathtt{N}^{\mathtt{w}}_{\psi}, \mathtt{S}^{\mathtt{w}}_{\psi} \rangle$, with identical number of nodes and labels, and *vice versa*.

It follows that:

- If $\mathtt{v}$ is a simple leaf, $e_{\varphi}(\overline{0}, \mathtt{v}) = k = e_{\psi}(\overline{0}, \mathtt{w})$, by item (i) ($k = 0$ or $k = 1$).
- If $\mathtt{v}$ is a looping leaf, then $\mathtt{v}$ and $\mathtt{w}$ belong to l-isomorphic loops and so are assigned the same equation system, by item (iii). Identical equation systems have identical solutions or no solutions. More in particular:
  - $e_{\varphi}(\alpha, \mathtt{v}) = (\mathbf{s_m} =_{\mathbf{U}} \mathbf{s_n})$ if and only if $e_{\psi}(\beta, \mathtt{w}) = (\mathbf{s_m} =_{\mathbf{U}} \mathbf{s_n})$, for some $\mathscr{L}_U$-terms s_m and s_n.
  - $e_{\varphi}(\alpha + m, \mathtt{v}) = k$ if and only if $e_{\psi}(\beta + n, \mathtt{w}) = k$, for some $m, n \in \omega$, if the equation system in question has a unique solution in V.
- If $\mathtt{v}$ is a simple point and $e_{\varphi}(\alpha, \mathtt{v})$ has a value, then one or more of the successors of $\mathtt{v}$ has already a value in $\text{Ł} \cup V$ at an ordinal $\gamma < \alpha$;[27] by item (ii), then, for some ordinal β the function $e_{\psi}(\beta, \mathtt{w})$ has the same value, for the very same reason.

[27] See, for example, item 1.1 in Definition 5.2.3.

- Finally, the evaluation functions move towards the root node either via paths of simple points or via loops that are closer to the root than those already evaluated. In both cases (as seen with the two previous bullet points), e_φ and e_ψ give identical values to nodes with identical labels.

$\square$

5.3.2 The canonical evaluation

Theorem 5.3.7 enables us to speak of the value of the *sentence* σ, rather than the value of a node labelled with σ in a graph, i.e. to define a *canonical evaluation*.

Definition 5.3.9. (Canonical evaluation for sentences)
Define the *canonical evaluation function* $\mathscr{C} : SENT_{\mathscr{L}_T^{\rightarrow}} \rightharpoonup \text{Ł} \cup V$ as follows:

$$\mathscr{C}(\varphi) :\simeq e_\varphi(\omega_1^{\mathrm{CK}}, \mathtt{r}).$$

The *Kleene equality* symbol "$\simeq$" indicates that either $\mathscr{C}(\varphi)$ and $e_\varphi(\omega_1^{\mathrm{CK}}, \mathtt{r})$ have a value and such value is the same, or that neither of them does. ▲

The canonical evaluation $\mathscr{C}$ is the model of $\mathscr{L}_T^{\rightarrow}$-sentences I propose in this chapter. The choice of the root of $\mathtt{S}_\varphi$ to define the canonical value of φ is only a matter of convention, since Theorem 5.3.7 entails that it could be replaced by any node labelled with φ. Let me now review the main properties of $\mathscr{C}$.

First, $\mathscr{C}$ enjoys the following closure properties, which codify the naïve rules for introduction and elimination of T, *modus ponens*, and a strong form of conditional introduction, together with the Kripkean rules for the remaining connectives and quantifiers of $\mathscr{L}_T^{\rightarrow}$.

Theorem 5.3.10. (Closure properties of $\mathscr{C}$)
For all $\varphi, \psi \in \mathscr{L}_T^{\rightarrow}$, $\chi(x) \in FOR_{\mathscr{L}_T^{\rightarrow}}$, the following hold:

- *If* $\mathscr{C}(\neg\varphi) = 1$*, then* $\mathscr{C}(\varphi) = 0$.

  If $\mathscr{C}(\varphi) = 0$*, then* $\mathscr{C}(\neg\varphi) = 1$.

- *If* $\mathscr{C}(\varphi \wedge \psi) = 1$*, then* $\mathscr{C}(\varphi) = 1$ *and* $\mathscr{C}(\psi) = 1$.

  If $\mathscr{C}(\varphi) = 1$ *and* $\mathscr{C}(\psi) = 1$*, then* $\mathscr{C}(\varphi \wedge \psi) = 1$.

- *If* $\mathscr{C}(\varphi \vee \psi) = 1$*, then* $\mathscr{C}(\varphi) = 1$ *or* $\mathscr{C}(\psi) = 1$.

  If $\mathscr{C}(\varphi) = 1$ *or* $\mathscr{C}(\psi) = 1$*, then* $\mathscr{C}(\varphi \vee \psi) = 1$.

- *If* $\mathscr{C}(\varphi \to \psi) = 1$ *and* $\mathscr{C}(\varphi) = 1$*, then* $\mathscr{C}(\psi) = 1$.

  If $\mathscr{C}(\varphi) = 0$*, or* $\mathscr{C}(\psi) = 1$*,*

  or $\mathscr{C}(\varphi) = j$*,* $\mathscr{C}(\psi) = k$*, and* $j \leq k$*, then* $\mathscr{C}(\varphi \to \psi) = 1$.

- *If* $\mathscr{C}(\forall x \chi(x)) = 1$*, then* $\mathscr{C}(\chi(t) = 1$ *for all* $t \in CTER_{\mathscr{L}_T^{\to}}$.

  If $\mathscr{C}(\chi(t)) = 1$ *for all* $t \in CTER_{\mathscr{L}_T^{\to}}$*, then* $\mathscr{C}(\forall x \chi(x)) = 1$.

- *If* $\mathscr{C}(\varphi) = 1$*, then* $\mathscr{C}(T\ulcorner\varphi\urcorner) = 1$.

  If $\mathscr{C}(T\ulcorner\varphi\urcorner) = 1$*, then* $\mathscr{C}(\varphi) = 1$.

Proof.

The proof is quite mechanical, I will show only the case of the conditional.

Let $\mathscr{C}(\varphi \to \psi) = 1, \mathscr{C}(\varphi) = 1$. Then $e_{\varphi\to\psi}(\omega_1^{\mathrm{CK}}, \mathtt{r}_1) = 1$ and $e_{\varphi}(\omega_1^{\mathrm{CK}}, \mathtt{r}_2) = 1$, for $l_{\varphi\to\psi}(\mathtt{r}_1) = \varphi \to \psi$, $l_{\varphi}(\mathtt{r}_2) = \varphi$. $\mathtt{r}_1$ is not a leaf ($\to$ is binary). Since ω_1^{CK} is the closure ordinal of $e_{\varphi\to\psi}(\alpha, \mathtt{v}_\mathtt{n})$, by Theorem 5.2.4 its evaluation clause is:

$$
\begin{aligned}
e_{\varphi\to\psi}(\omega_1^{\mathrm{CK}}, \mathtt{r}_1) = 1 \text{ iff } & e_{\varphi\to\psi}(\omega_1^{\mathrm{CK}}, \mathtt{v}_1) = 0, \text{ or} \\
& e_{\varphi\to\psi}(\omega_1^{\mathrm{CK}}, \mathtt{v}_2) = 1, \text{ or} \\
& e_{\varphi\to\psi}(\omega_1^{\mathrm{CK}}, \mathtt{v}_1) = j, e_{\varphi\to\psi}(\omega_1^{\mathrm{CK}}, \mathtt{v}_2) = k \text{ and } j \leq k
\end{aligned}
\tag{5.1}
$$

for $l_{\varphi\to\psi}(\mathtt{v}_1) = \varphi$, $l_{\varphi\to\psi}(\mathtt{v}_2) = \psi$. By Definition 5.1.2 $\mathtt{v}_1, \mathtt{v}_2 \in \mathtt{N}_{\varphi\to\psi}$ (they are successors of $\mathtt{r}_1$, since the latter is a root node). Since $e_{\varphi}(\omega_1^{\mathrm{CK}}, \mathtt{r}_2) = 1$ and $l_{\varphi}(\mathtt{r}_2) =$

$l_{\varphi\to\psi}(\mathtt{v}_1)$, by Theorem 5.3.7 also $e_{\varphi\to\psi}(\omega_1^{\mathrm{CK}}, \mathtt{v}_1) = 1$ (and so $e_{\varphi\to\psi}(\omega_1^{\mathrm{CK}}, \mathtt{v}_1) \neq 0$, by Theorem 5.2.4). By equation (5.1), $e_{\varphi\to\psi}(\omega_1^{\mathrm{CK}}, \mathtt{v}_2) = k$ and $1 \leq k$, but $k \in V$, so $k = 1$. Applying again Theorem 5.3.7, we have that $e_{\psi}(\omega_1^{\mathrm{CK}}, \mathtt{r}_3) = 1$, for $l_{\psi}(\mathtt{r}_3) = l_{\varphi\to\psi}(\mathtt{v}_2) = \psi$, and by definition $\mathscr{C}(\psi) = 1$, as desired.

Suppose now that $\mathscr{C}(\varphi) = 0$, or $\mathscr{C}(\psi) = 1$, or ($\mathscr{C}(\varphi) = j$, $\mathscr{C}(\psi) = k$, (for $j \leq k$). Then, $e_{\varphi}(\omega_1^{\mathrm{CK}}, \mathtt{r}_2) = 0$, or $e_{\psi}(\omega_1^{\mathrm{CK}}, \mathtt{r}_3) = 1$, or $e_{\varphi}(\omega_1^{\mathrm{CK}}, \mathtt{r}_2) = j$ and $e_{\psi}(\omega_1^{\mathrm{CK}}, \mathtt{r}_3) = k$ (using the same labels as above). Let the last case be given (otherwise it is similar). By Theorem 5.3.7, $e_{\varphi\to\psi}(\omega_1^{\mathrm{CK}}, \mathtt{v}_1) = j$ and $e_{\varphi\to\psi}(\omega_1^{\mathrm{CK}}, \mathtt{v}_2) = k$ ($\mathtt{v}_1$ and $\mathtt{v}_2$ exist by Definition 5.1.2), and by equation (5.1) $e_{\varphi\to\psi}(\omega_1^{\mathrm{CK}}, \mathtt{r}_1) = 1$, i.e. $\mathscr{C}(\varphi \to \psi) = 1$, as desired. $\square$

This result will be used in Section 5.5 to associate (a family of) notions of logical consequence with the present construction, which will yield a genuine conditional introduction rule that improves on that of the ŁK-construction.[28] Before being able to do that, however, we will need the technical machinery of Section 5.4 (in particular Theorem 5.4.2), which has also independent interest and purposes.

The next result generalizes Theorem 5.3.10 to the entire value space $Ł \cup V$, giving a fuller picture of the partial interpretation of $\mathscr{L}_T^{\to}$-sentences offered by $\mathscr{C}$.

Theorem 5.3.11. (Behavior of $\mathscr{C}$)[29]

For every $\varphi \in \mathscr{L}_T^{\to}$, the following hold (I use s_ψ indicate u_ψ or a $\mathscr{L}_U$-constant).

(A) *If φ is an atomic arithmetical sentence, then $\mathscr{C}(\varphi)$ is 1, if $\mathbb{N} \models \varphi$; 0, if $\mathbb{N} \not\models \varphi$.*

$(\neg)$ *If φ is $\neg\psi$, then $\mathscr{C}(\varphi)$ is:*

- $1 - \mathscr{C}(\psi)$, *if $\mathscr{C}(\psi) \in V$,*
- $\mathbf{u}_{\neg\psi} =_{\mathbf{U}} \mathbf{1} - \mathbf{u}_\psi$, *if $\mathscr{C}(\psi) \in Ł$.*

[28] For the *genuineness* condition, see Definition 2.7.1 in Section 2.7. Genuine conditions for conditional introduction were discussed in the context of Field's theory in Subsection 3.5.2 and in the context of the ŁK-construction in Subsection 4.4.3.

[29] This theorem can be proven in a way that is analogous to the proof of Theorem 5.3.10.

($\wedge$) *If φ is $\psi \wedge \chi$, then $\mathscr{C}(\varphi)$ is:*

- 0, *if* $\mathscr{C}(\psi) = 0$ *or* $\mathscr{C}(\chi) = 0$.
- $\min(\mathscr{C}(\psi), \mathscr{C}(\chi))$, *if* $\mathscr{C}(\psi), \mathscr{C}(\chi) \in V$.
- $\mathbf{u}_{\psi\wedge\chi} =_{\mathbf{U}} \min(\mathbf{s}_{\psi}, \mathbf{s}_{\chi})$, *if:*
  - $\mathscr{C}(\psi), \mathscr{C}(\chi) \in$ Ł, *or*
  - $\mathscr{C}(\psi) \in$ Ł, $\mathscr{C}(\chi) \in V \setminus \{0\}$, *or*
  - $\mathscr{C}(\chi) \in$ Ł, $\mathscr{C}(\psi) \in V \setminus \{0\}$.

($\vee$) *If φ is $\psi \vee \chi$, then $\mathscr{C}(\varphi)$ is:*

- 1, *if* $\mathscr{C}(\psi) = 1$ *or* $\mathscr{C}(\chi) = 1$.
- $\max(\mathscr{C}(\psi), \mathscr{C}(\chi))$, *if* $\mathscr{C}(\psi), \mathscr{C}(\chi) \in V$.
- $\mathbf{u}_{\psi\vee\chi} =_{\mathbf{U}} \max(\mathbf{s}_{\psi}, \mathbf{s}_{\chi})$, *if:*
  - $\mathscr{C}(\psi), \mathscr{C}(\chi) \in$ Ł, *or*
  - $\mathscr{C}(\psi) \in$ Ł, $\mathscr{C}(\chi) \in V \setminus \{1\}$, *or*
  - $\mathscr{C}(\chi) \in$ Ł, $\mathscr{C}(\psi) \in V \setminus \{1\}$.

($\rightarrow$) *If φ is $\psi \rightarrow \chi$, then $\mathscr{C}(\varphi)$ is:*

- 1, *if* $\mathscr{C}(\psi) = 0$ *or* $\mathscr{C}(\chi) = 1$.
- $\min[1, (1 - \mathscr{C}(\psi) + \mathscr{C}(\chi))]$, *if* $\mathscr{C}(\psi), \mathscr{C}(\chi) \in V$.
- $\mathbf{u}_{\psi\rightarrow\chi} =_{\mathbf{U}} \min[\mathbf{1}, (\mathbf{1} - \mathbf{s}_{\psi} + \mathbf{s}_{\chi})]$, *if:*
  - $\mathscr{C}(\psi), \mathscr{C}(\chi) \in$ Ł, *or*
  - $\mathscr{C}(\psi) \in$ Ł, $\mathscr{C}(\chi) \in V \setminus \{1\}$, *or*
  - $\mathscr{C}(\chi) \in$ Ł, $\mathscr{C}(\psi) \in V \setminus \{0\}$.

($\forall$) *If φ is $\forall x \chi(x)$, then $\mathscr{C}(\varphi)$ is:*

- 0, *if* $\mathscr{C}(\chi(en(k))) = 0$ *for some* $k \in \omega$.

- $\inf\{\mathscr{C}(\chi(en(k)))\,|\,k\in\omega\}$, *if* $\mathscr{C}(\chi(en(k)))\in V$ *for all* $k\in\omega$.
- $\mathbf{u}_{\forall\mathbf{x}\chi(\mathbf{x})} =_{\mathbf{U}} \inf\{\mathbf{s}_{\chi(\mathbf{en}(\mathbf{k}))}|\mathbf{k}\in\omega\}$ *if* $\mathscr{C}(\chi(en(k)))\in Ł\cup V\setminus\{0\}$ *for all* $k\in\omega$ *and* $\mathscr{C}(\chi(en(n)))\in Ł$ *for some* $n\in\omega$.

(T) *If* φ *is* $T\ulcorner\psi\urcorner$, *then* $\mathscr{C}(\varphi)$ *is:*

- $\mathscr{C}(\psi)$, *if* $\mathscr{C}(\psi)\in V$.
- $\mathbf{u}_{\mathbf{T}\ulcorner\psi\urcorner} =_{\mathbf{U}} \mathbf{u}_{\psi}$, *if* $\mathscr{C}(\psi)\in Ł$.

Note that, by Theorem 5.2.4 and partiality, at most one but not at least one condition per bullet point is satisfied.

Moreover, the truth predicate characterized by $\mathscr{C}$ respects the intersubstitutivity of truth for every sentence that receives a value.

Corollary 5.3.12. (Intersubstitutivity of truth (formulation by Field))[30]
For every $\varphi,\psi,\chi\in\mathscr{L}_T^{\rightarrow}$, *if* ψ *and* χ *are alike except that* ψ *has an occurrence of* φ *of* $\mathscr{L}_T$ *where* χ *has an occurrence of* $T\ulcorner\varphi\urcorner$, *then*

$$\mathscr{C}(\psi)=k \quad \textit{if and only if} \quad \mathscr{C}(\chi)=k \qquad (\textit{for } k\in V).$$

An analogous result holds if $\mathscr{C}(\varphi)\in Ł$ *(*modulo *addition or removal of equations defined using the identity function).*

Finally we have another corollary, whose relevance comes from all the cases in which $\mathscr{C}$ succeeds in assigning a numerical value to $\mathscr{L}_T^{\rightarrow}$-sentences.

Corollary 5.3.13. (Tarski biconditionals for sentences that get a value in V)
For every $\varphi\in\mathscr{L}_T^{\rightarrow}$, *the following holds:*

$$\textit{there is a } k\in V \textit{ s.t. } \mathscr{C}(\varphi)=k \quad \textit{if and only if} \quad \mathscr{C}(\varphi\leftrightarrow T\ulcorner\varphi\urcorner)=1.$$

[30]The proof is an application of Theorem 5.3.11, along the lines of the proof of Corollary 2.4.2.

This result completes a quick survey of the main facts on the canonical evaluation. In the next chapter, I will argue for the interpretation of $\mathscr{L}_T^{\rightarrow}$-sentences that the canonical evaluation provides, arguing that it constitutes a satisfactory way to address FIELD'S PROGRAM and to handle truth-theoretic paradoxes. In the next two sections, I will explore the treatment of determinateness that emerges from the present model and a notion of logical consequence associated with it.

5.4 Determinateness and revenge

We have seen that revenge phenomena make it inconsistent for Field's theory (a total theory) to have a bivalent, limit, or idempotent determinateness operator.[31] The situation is somehow different in partial theories: while Kripke's theory cannot express an operator that exceeds the power of strong Kleene negation, ŁK-construction yields a better notion of determinateness that, however, remains quite weak.[32] The semantics provided by the canonical evaluation, instead, yields a unique, partial but strong notion of determinateness, that behaves bivalently and idempotently on all the sentences that are assigned a numerical value, and that is a natural limit of a converging sequence of iterations of a single definable operator. Let me recall the definition of *Fieldian determinateness*, or *determinateness* for short, given in Lemma 4.3.6.

Definition 5.4.1. (Fieldian determinateness operator)

For every $\varphi \in \mathscr{L}_T^{\rightarrow}$, put $\mathcal{D}(\varphi) := \neg(\varphi \rightarrow \neg\varphi)$ and call $\mathcal{D}$ a *Fieldian determinateness operator*, or *determinateness operator*, for short.

Let $\mathcal{D}\mathcal{D}(\varphi)$ abbreviate $\mathcal{D}(\mathcal{D}(\varphi))$, and let $\mathcal{D}^{\underline{n}}(\varphi)$ be a string of n iterations of $\mathcal{D}$ applied to φ. More in general, let Nt be a univalent recursive ordinal notation system, whose range is O_{Nt}.[33] Define a hierarchy of determinateness operators *à la* Field

[31] See the discussion in Subsection 3.6.2.
[32] See the discussion at the end of Subsection 4.3.
[33] For ordinal notation systems, see [Rogers 1987], Chapter 11.7

for ordinals in O_{Nt}.[34] As for ω iterations of $\mathcal{D}$, let r be the primitive recursive function from positive integers and sentences to sentences s.t. $r(n, \ulcorner\varphi\urcorner) := \mathcal{D}^{\underline{n}}(\varphi)$, and define $\mathcal{D}^{\underline{\omega}}(\varphi)$ as $\forall n T \ulcorner \dot{r}(n, \ulcorner\varphi\urcorner)\urcorner$. ▲

I collect in one theorem all the fundamental results about the treatment of determinateness in the canonical evaluation.

Theorem 5.4.2. (Partial, Bivalent, Limit, and Idempotent Determinateness) *For every* $\varphi \in \mathcal{L}_T^{\rightarrow}$ *such that* $\mathscr{C}(\varphi) = k$*, for* $k \in V$*, the following holds.*

1. *For all ordinals* $\alpha \in O_{Nt}$*,* $\mathscr{C}(\mathcal{D}^{\underline{\alpha}}(\varphi)) \in V$*. In particular:*

$$\mathscr{C}(\mathcal{D}^{\underline{\alpha+1}}(\varphi)) = 1 - \min[1, (1 - \mathscr{C}(\mathcal{D}^{\underline{\alpha}}(\varphi)) + 1 - \mathscr{C}(\mathcal{D}^{\underline{\alpha}}(\varphi)))]. \tag{5.2}$$

$$\mathscr{C}(\mathcal{D}^{\underline{\delta}}(\varphi)) = \inf\{\mathscr{C}(\mathcal{D}^{\underline{\gamma}}(\varphi)) | \gamma < \delta\}, \text{ if } \delta \text{ is a limit ordinal.} \tag{5.3}$$

2. *For each numerical Łukasiewicz value space* V *(see Subsection 5.2.1), there is a unique ordinal* $\delta^V \in O_{Nt}$ *such that, for all* $\delta \in O_{Nt}$ *such that* $\delta \geq \delta^V$*:*

$$\begin{aligned} &\mathscr{C}(\mathcal{D}^{\underline{\delta}}(\varphi)) = 1 \;\; \textit{if and only if} \;\; \mathscr{C}(\varphi) = 1. \\ &\mathscr{C}(\mathcal{D}^{\underline{\delta}}(\varphi)) = 0 \;\; \textit{if and only if} \;\; \mathscr{C}(\varphi) \in V \textit{ and } \mathscr{C}(\varphi) \neq 1. \end{aligned} \tag{5.4}$$

Proof.
Ad 1, let $\alpha \in O_{Nt}$ and reason by induction. By definition, $\mathcal{D}^{\underline{\alpha+1}}(\varphi) = \neg(\mathcal{D}^{\underline{\alpha}}(\varphi) \rightarrow \neg\mathcal{D}^{\underline{\alpha}}(\varphi))$. By IH, $\mathscr{C}(\mathcal{D}^{\underline{\alpha}}(\varphi)) \in V$. Applying Theorem 5.3.11 multiple times we get $\mathscr{C}(\neg(\mathcal{D}^{\underline{\alpha}}(\varphi) \rightarrow \neg\mathcal{D}^{\underline{\alpha}}(\varphi))) \in V$, i.e. $\mathscr{C}(\mathcal{D}^{\underline{\alpha+1}}(\varphi)) \in V$, while equation (5.2) holds by construction. The reasoning for limit ordinals and equation (5.3) is similar.

[34] See Subsection 3.6.1. This formulation is vague, but making it precise would require an ultimately unnecessary detour through the theory of ordinal notations. The iterations of $\mathcal{D}$ allowed given my condition on Nt are *much* shorter than those of [Field 2008], but neither longer iterations nor stronger notations are needed, as the proof of Theorem 5.4.2 shows.

Ad 2, note that if $\mathscr{C}(\varphi) \leq 1/2$, then $\mathscr{C}(\mathcal{D}(\varphi)) = 0$ (by Theorem 5.3.11 and simple calculation). Let $V = \{0, 1/2m, \ldots, 2m-1/2m, 1\}$. If $\mathscr{C}(\varphi) = k/2m$ (for $1/2 < k/2m < 1$), then (again by Theorem 5.3.11 and simple calculation):

$$\mathscr{C}(\mathcal{D}(\varphi)) = 1 - \min[1, (1 - k/2m + (1 - k/2m))] = (2k-2m)/2m. \tag{5.5}$$

So, if $\mathscr{C}(\varphi) = k/2m$ (for $1/2 < k/2m < 1$), then $\mathscr{C}(\mathcal{D}^{\underline{m}}(\varphi)) = 0$ applying at most m times equation (5.5). For all $\alpha \in O_{Nt}$ s.t. $\alpha \geq m$ (by item 1 of the theorem, (5.5), and Theorem 5.3.11), $\mathscr{C}(\mathcal{D}^{\underline{\alpha}}(\varphi)) = 0$. This proves item 2 for every finite V, in which case δ^V is the smallest $l \leq m$ s.t. $\mathscr{C}(\mathcal{D}^{\underline{l}}(\varphi)) = 0$ (uniqueness is immediate). Let $V = [0,1]$ and $\mathscr{C}(\varphi) = k$, for $0 \leq k < 1$. So, there is a $j/2n$ s.t. $k < j/2n < 1$. By the previous result on finite Vs, $\mathcal{D}^{\underline{n}}(\varphi) = 0$. Since such a $\mathcal{D}^{\underline{n}}$ exists for each $k \in V$, the claim holds also for $\mathcal{D}^{\underline{\omega}}$ (which exists and is unique by item 1 of the theorem). As above, for all $\delta \in O_{Nt}$ s.t. $\delta \geq \omega$ (by item 1, equation (5.5), and Theorem 5.3.11), $\mathscr{C}(\mathcal{D}^{\underline{\delta}}(\varphi)) = 0$, and this shows that $\delta^V = \omega$, if $V = [0,1]$. □

Unlike in Field's theory, in the semantics provided by the canonical evaluation there is no never-ending and never-converging hierarchy of stronger and stronger determinateness operators.[35] The hierarchy of determinateness operators we have here converges to a single operator $\mathcal{D}^{\underline{\delta^V}}$ at a small ordinal, depending on $|V|$. For each V, the operator $\mathcal{D}^{\underline{\delta^V}}$ is *bivalent*, *limit*, and *idempotent* on every $\varphi \in \mathscr{L}_T^{\rightarrow}$ such that $\mathscr{C}(\varphi) \in V$, and it behaves as described by (5.2) and (5.3), even if φ features some occurrences of $D^{\underline{\beta}}$, for $\beta \in O_{Nt}$, possibly $\beta > \delta^V$. So, we can take $\mathcal{D}^{\underline{\delta^V}}$ as *the* notion of determinateness given by the present model, for each choice of V.[36]

The treatment of determinateness offered by the canonical evaluation *approximates* the ideal notion of *total*, bivalent, limit, and idempotent determinateness –

[35] See Subsection 3.6.1.

[36] $\mathcal{D}^{\underline{\delta^V}}$ is unique *modulo* the choice of Nt, but the ordinals involved are so small (ω at most) that there is basically no dependence on the notation adopted.

such ideal is only expressible in the meta-theory – in the sense of Subsection 3.6.2. For each choice of V, the operator $\mathcal{D}^{\underline{\delta V}}$ approximates the meta-theoretic ideal as much as possible, in line with the criteria for assigning numerical values that are adopted by the canonical evaluation. Such approximations improve in their coverage of cases as the cardinality of V increases. But here, unlike in Field's theory, for each choice of V we have *one* well-defined approximating notion, *one* well-defined set of covered cases, and *one* well-defined set of non-covered residuals.

The non-covered residuals are refined into two categories: the sentences that are given a value in Ł, and the sentences that are given no value. Let me focus on the first possibility. The case of determinateness, amongst others, demonstrates the usefulness of assigning Łukasiewicz equations as semantic values. For example, for every choice of V, interpreting connectives and quantifiers *à la* Łukasiewicz forces partiality if weak naïveté is adopted (Theorem 4.1.2). For the case of $V = [0,1]$, this was (implicitly) shown by Restall, and a proof using determinateness was given by Field.[37] We can reproduce Field's proof using the canonical evaluation: the sentence used by Field to prove Restall's result, that features the determinateness operator, generates a semantic graph that has a value in Ł but not in $[0,1]$, and thus in no V. Here, thus, a set of Łukasiewicz equations, used as semantic values, shows explicitly why and how we cannot have a *total* notion of determinateness that behaves as in Theorem 5.4.2.[38] I state again the second half of Theorem 4.1.2.

Theorem 5.4.3. (Restall)

There is no total function $\mathscr{E}^{\infty}_{Ł} : SENT_{\mathscr{L}^{\rightarrow}_{T}} \longmapsto [0,1]$ *s.t.:*

1. $\mathscr{E}^{\infty}_{Ł}$ *agrees with* $\mathbb{N}$ *on atomic arithmetical sentences,*

2. *For every* $\varphi \in \mathscr{L}^{\rightarrow}_{T}$, $\mathscr{E}^{\infty}_{Ł}(\varphi) = \mathscr{E}^{\infty}_{Ł}(T\ulcorner\varphi\urcorner)$,

[37] See [Restall 1992] and [Field 2008], p. 92.

[38] I will argue more in general for the philosophical significance of Łukasiewicz equations as semantic values in Section 6.3.

3. $\mathscr{E}^{\infty}_{\text{Ł}}$ is a $[0,1]$-valued Łukasiewicz evaluation function.

Proof (strategy by Field).

Let ρ be the fixed point of $\neg\forall n T^{\ulcorner} \underset{\cdot}{r}(n,x)^{\urcorner}$.[39] Call ρ *Restall's sentence*; this sentence is clearly a variant of McGee's sentence. Consider the following graph.

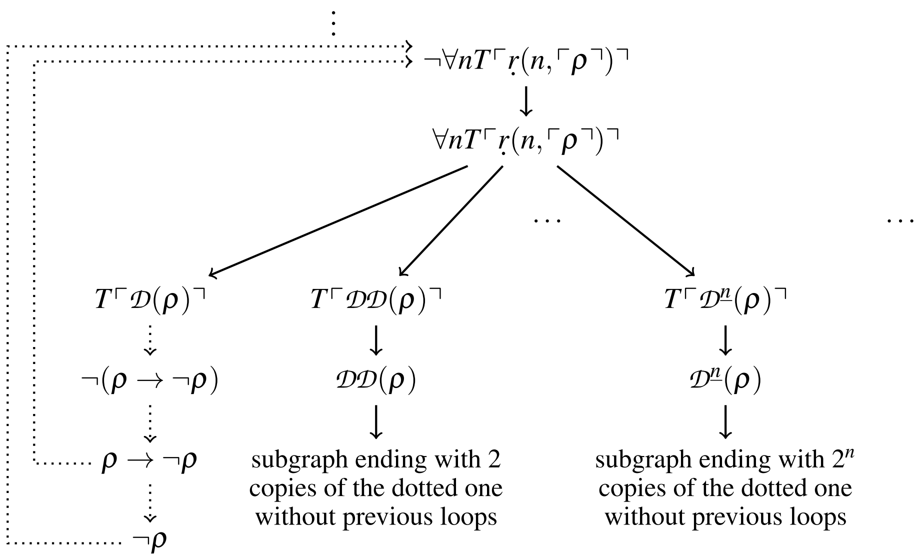


Figure 5.9: The semantic graph of Restall's sentence

The evaluation e_ρ associates with all the nodes of $\langle \mathtt{N}_\rho, \mathtt{S}_\rho \rangle$ exactly one infinite system of equations, ending at level $\omega+1$. Suppose that this system has exactly one solution in $[0,1]$, where connectives and quantifiers are interpreted *à la* Łukasiewicz and truth is interpreted following weak naïveté. So, we can suppose that the system has a unique solution in the canonical evaluation.

Suppose that $\mathscr{C}(\rho) = 1$. An easy induction shows that the equation associated with $\mathcal{D}^{\underline{n}}(\rho)$, for all $n \in \omega$, has 1 as its only solution. So, $\mathscr{C}(\forall n T^{\ulcorner}\underset{\cdot}{r}(n,{}^{\ulcorner}\rho^{\urcorner})^{\urcorner}) = 1$ and $\mathscr{C}(\neg\forall n T^{\ulcorner}\underset{\cdot}{r}(n,{}^{\ulcorner}\rho^{\urcorner})^{\urcorner}) = 0$. But this sentence is ρ, so this assignment is wrong.

Suppose that $\mathscr{C}(\rho) = k$, for $k \in [0,1]$ and $k < 1$. If $0 \leq k \leq 1/2$, the only so-

[39] Recall that the function r was defined in Definition 5.4.1 as the primitive recursive function from positive integers and sentences to sentences s.t. $r(n, {}^{\ulcorner}\varphi^{\urcorner}) := \mathcal{D}^{\underline{n}}(\varphi)$.

lution to the equation associated with each sentence of the form $\mathcal{D}^{\underline{n}}(\rho)$ is 0, and so $\mathscr{C}(\forall n T^{\ulcorner} \underset{.}{r}(n, {}^{\ulcorner}\rho^{\urcorner})^{\urcorner}) = 0$ and $\mathscr{C}(\neg\forall n T^{\ulcorner} \underset{.}{r}(n, {}^{\ulcorner}\rho^{\urcorner})^{\urcorner}) = 1$, which is ρ, against our supposition. If $1/2 \leq k < 1$ then, by the proof of Theorem 5.4.2, there is a $j \in \omega$ s.t. the only possible solution associated with $\mathcal{D}^{\underline{j}}(\rho)$ is $\leq 1/2$. Then consider $\mathcal{D}^{\underline{j+1}}(\rho)$ and reason as above. $\square$

5.5 Logical consequence and conditional introduction

In this section, I will generalize the canonical evaluation to *many*, uniformly defined evaluations that share the fundamental features of $\mathscr{C}$. This will make it possible to define a family of semantic notions of logical consequence, generalizing Theorem 5.3.10 and turning the closure properties given there in introduction and elimination rules, including a genuine conditional introduction.

The basic idea is similar to what we have in Field's theory, in Kripke's theory, or in the ŁK-construction. Suppose that we have a set of functions defined as in Definition 5.3.9, with the possible difference of using non-minimal fixed points of Definition 5.2.3 (I will be more specific in the following): call any such function $\mathscr{C}'$ and the set $\mathcal{Q}$. We can then define a notion of logical consequence thus:

$$\Gamma \models^{\mathcal{Q}}_{\mathscr{C}} \psi \text{ if and only if for every function } \mathscr{C}' \in \mathcal{Q}, \text{ if for every } \varphi \in \Gamma, \mathscr{C}'(\varphi) = 1, \text{ then also } \mathscr{C}'(\psi) = 1.$$

To make this definition precise, we must specify the functions $\mathscr{C}'$.

Now, the canonical evaluation is defined using the evaluations of nodes in semantic graphs. The latter, in turn, can be seen as defined inductively: for every $\varphi \in \mathscr{L}_T^{\rightarrow}$, e_φ is the *least* fixed point of the positive elementary formula that results applying Definition 5.2.3 to $\mathtt{N}_\varphi$.[40] For every $\varphi \in \mathscr{L}_T^{\rightarrow}$, however, there are also non-

[40]Such formulae are inductive in other sets too (see the proof of Theorem 5.2.4, item 1), but this is

minimal fixed points of Definition 5.2.3 applied to $\mathtt{N}_\varphi$. Denote a generic, possibly non-minimal fixed point of Definition 5.2.3 applied to $\mathtt{N}_\varphi$ with e^+_φ. A specification is in order: many different functions are fixed points of Definition 5.2.3 applied to $\mathtt{N}_\varphi$, including functions that have domain or range that strictly extend those of e_φ. Such functions, however, contain information that is not pertinent to evaluate the nodes of $\mathtt{N}_\varphi$ and should therefore be excluded from our consideration. So, from now on, I will only consider fixed points e^+_φ having the same domain and range as e_φ, even though it may be that $e_\varphi \subsetneq e^+_\varphi$.[41]

In general, the situation is analogous to what we have in any theory employing inductive constructions. Non-minimal fixed points of Definition 5.2.3, however, are not comparable to non-minimal Kripke fixed points in terms of their results. To see this, take τ. The truth-teller is in the truth-set of some consistent non-minimal Kripke fixed points and in the falsity-set of others.[42] However, for every φ s.t. there is a node labelled as τ in $\mathtt{N}_\varphi$, there are no fixed points e^+_φ extending e_φ that give a value different from $\mathbf{u}_\tau =_{\mathbf{U}} \mathbf{u}_\tau$ to any node labelled with τ in $\mathtt{N}_\varphi$. In other words, τ *must* have value $\mathbf{u}_\tau =_{\mathbf{U}} \mathbf{u}_\tau$ in the present semantics, no non-minimal fixed point of Definition 5.2.3 can change this. The reason is that every non-minimal fixed point of Definition 5.2.3 must respect the condition that, if a value is assigned to τ, it is the unique solution of $\mathbf{u}_\tau =_{\mathbf{U}} \mathbf{u}_\tau$, which is impossible.

Another difference with Kripke's theory, or the ŁK-construction, is that I did not define inductively some sets, associating an evaluation function with them only later: Definition 5.2.3 builds inductively the evaluation functions themselves. So, I will only consider non-minimal fixed points of Definition 5.2.3 that are *functions*.

not relevant for our concerns. Even if it is common to speak of fixed points of *monotone operators*, I did not associate explicitly such an operator with Definition 5.2.3, for readability; using the well-known duality between positive elementary formulae and monotone operators (see [Moschovakis 1974], [Pohlers 2009]), I will simply speak of fixed points of Definition 5.2.3 (applied to some $\mathtt{N}_\varphi$).

[41] For every $\varphi \in \mathscr{L}_T^{\rightarrow}$, the domain of e_φ is $Ord \times \mathtt{N}_\varphi$ – it is not the subset of $Ord \times \mathtt{N}_\varphi$ on which e_φ yields a value. Notice that, for every partial e_φ, there are non-minimal e^+_φ that are total functions.

[42] For this reason, there are variants of Field's theory that assign value $\mathbf{1}$ to τ and consistent non-minimal fixed points of Υ assigning value 1 to τ as well.

Thus no question of consistency of a fixed point arises.

Finally, recall that Theorem 5.3.7 ensures that two evaluations e_φ and e_ψ behave uniformly on nodes (in $\mathtt{N}_\varphi$ and $\mathtt{N}_\psi$ respectively) that have the same label: if $\mathtt{v} \in \mathtt{N}_\varphi$, $\mathtt{w} \in \mathtt{N}_\psi$, and $l_\varphi(\mathtt{v}) = l_\psi(\mathtt{w})$, either e_φ and e_ψ assign the same value to $\mathtt{v}$ and $\mathtt{w}$ (respectively), or both assign no value to them. This, however, is no longer guaranteed for non-minimal evaluations e^+_φ and e^+_ψ that assign values to nodes (in $\mathtt{N}_\varphi$ and $\mathtt{N}_\psi$ respectively) to which e_φ or e_ψ fail to assign a value. So, in order to define suitable generalizations of the canonical evaluation, we must make sure that the uniformity of the minimal evaluations of nodes is inherited by the non-minimal ones. I will make these requirements precise in the following definition.

Definition 5.5.1. (Some notions useful to generalize the canonical evaluation)

- For every $\varphi \in \mathscr{L}^{\rightarrow}_T$, let $\mathfrak{G}_\varphi$ be a set of fixed points of Definition 5.2.3 applied to $\mathtt{N}_\varphi$.[43] I will use $\mathcal{G}$ to indicate a set of sets of fixed points of Definition 5.2.3, so that $\mathcal{G}$ has the form $\{\mathfrak{G}_\varphi, \mathfrak{G}'_\varphi, \ldots, \mathfrak{G}_\psi, \mathfrak{G}'_\psi, \ldots, \mathfrak{G}_\chi, \mathfrak{G}'_\chi, \ldots\}$.
- Let $\mathcal{G}$ be a set of sets of fixed points of Definition 5.2.3; then, I will say that the set $Gen(\mathcal{G}) := \{\varphi \in \mathscr{L}^{\rightarrow}_T \mid \mathfrak{G}_\varphi \in \mathcal{G}\}$ is the set of its *generators*.[44]
- A set $\mathcal{G}$ of sets of fixed points of Definition 5.2.3 is *coherent* if, whenever two nodes $\mathtt{v}, \mathtt{w} \in \bigcup_{\varphi \in Gen(\mathcal{G})} \mathtt{N}_\varphi$ have the same label, for every $e^+_\psi, e^+_\chi \in \bigcup \mathcal{G}$ such that $\mathtt{v} \in \mathtt{N}_\psi$ and $\mathtt{w} \in \mathtt{N}_\chi$, we have that $e^+_\psi(\omega_1^{\mathrm{CK}}, \mathtt{v}) \simeq e^+_\chi(\omega_1^{\mathrm{CK}}, \mathtt{w})$. More intuitively, coherent sets of sets of fixed points of Definition 5.2.3 only contain

[43] By the uniqueness of semantic graphs and our restriction to non-minimal fixed points having the same domain and range as the minimal ones, for every $\varphi \in \mathscr{L}^{\rightarrow}_T$, we are excluding functions that are, at the same time, fixed points of Definition 5.2.3 applied to $\mathtt{N}_\varphi$ and fixed points of Definition 5.2.3 applied to $\mathtt{N}_\psi$ for ψ distinct from φ; so, the notation "$\mathfrak{G}_\varphi$" is justified. Note that fixed points of Definition 5.2.3 can be evaluation functions that do not assign any value to the nodes in their domains. Two such completely undefined evaluations obtained applying Definition 5.2.3 to two distinct sentences, however, are still distinct, as they have distinct domains.

[44] By our restriction to non-minimal fixed points having the same domain and range as the minimal ones, for every $\varphi \in \mathscr{L}^{\rightarrow}_T$, every set of sets of fixed points of Definition 5.2.3 applied to $\mathtt{N}_\varphi$ has exactly one generator, namely φ (see also the previous footnote), so for every set $\mathcal{G}$ of sets of fixed points of Definition 5.2.3 the set $Gen(\mathcal{G})$ is well-defined.

evaluations that are uniform as described in Theorem 5.3.7 (at ω_1^{CK}).

- A set $\mathcal{G}$ is *exhaustive* if $Gen(\mathcal{G}) = SENT_{\mathscr{L}_T^{\rightarrow}}$.

▲

Lemma 5.5.2.

1. *There are coherent and exhaustive sets of sets of fixed points of Definition 5.2.3.*

2. *If $\mathcal{G}$ is a coherent set of sets of fixed points of Definition 5.2.3, for every $\varphi \in Gen(\mathcal{G})$,* at the closure ordinal *there is exactly one $e_\varphi^+ \in \bigcup \mathcal{G}$. So, for every $\varphi \in Gen(\mathcal{G})$, if $e'_\varphi, e''_\varphi \in \bigcup \mathcal{G}$, for every $\mathtt{v} \in \mathtt{N}_\varphi$, we have $e'_\varphi(\omega_1^{\mathrm{CK}}, \mathtt{v}) \simeq e''_\varphi(\omega_1^{\mathrm{CK}}, \mathtt{v})$.*

Proof.
Ad 1, consider the evaluations $e_\varphi^\star$ that are defined exactly as the minimal evaluations e_φ (for every $\varphi \in \mathscr{L}_T^{\rightarrow}$), with the only exception that, at stage $\overline{0}$, they give value $1/2$ to every node in their domain that is labelled as a Yablo sentence. This extra clause does not conflict with any other clause of Definition 5.2.3, as no node labelled with a Yablo sentence receives a value in any minimal evaluation e_φ. Note that, for every $\varphi \in \mathscr{L}_T^{\rightarrow}$, the resulting function $e_\varphi^\star$ is a fixed point of Definition 5.2.3 applied to $\mathtt{N}_\varphi$. The set of all the evaluations $e_\varphi^\star$, for every $\varphi \in \mathscr{L}_T^{\rightarrow}$, is easily seen to be coherent and exhaustive, as it differs from the set of the minimal evaluations in treating every node labelled with a Yablo sentence as a $1/2$-valued leaf – every other node (possibly labelled with sentences having the now-evaluated Yablo sentences as subsentences) is treated in the usual compositional, bot-top way.[45]
Item 2 is immediate by the definition of coherence. □

[45] We know from CAB fixed points of Subsection 4.5.1 that assigning value $1/2$ to Yablo sentences is consistent with weak naïveté and our partial version of Łukasiewicz 3-valued semantics (and thus with weak naïveté and our partial versions of every Łukasiewicz semantics).

Thanks to this lemma, we can use coherent and exhaustive sets of sets of fixed points of Definition 5.2.3 to give generalized versions of the canonical evaluation.

Definition 5.5.3. (Quasi-canonical evaluations)

For every coherent and exhaustive set $\mathcal{G}$ of sets of fixed points of Definition 5.2.3, the *quasi-canonical evaluation induced by* $\mathcal{G}$ is the unique (possibly total) function $\mathscr{C}_{\mathcal{G}} : SENT_{\mathscr{L}_T^{\rightarrow}} \rightharpoonup \text{Ł} \cup V$ such that for every $\varphi \in \mathscr{L}_T^{\rightarrow}$:

$$\mathscr{C}_{\mathcal{G}}(\varphi) :\simeq e^{+}_{\varphi}(\omega_1^{\mathrm{CK}}, \mathtt{r}),$$

for every evaluation $e^{+}_{\varphi} \in \bigcup \mathcal{G}$ such that $\varphi \in Gen(\mathcal{G})$. ▲

For every coherent and exhaustive set $\mathcal{G}$ of sets of fixed points of Definition 5.2.3, existence and uniqueness of $\mathscr{C}_{\mathcal{G}}$ are immediate from Lemma 5.5.2.

Of course, the canonical evaluation is a quasi-canonical evaluation: it is the quasi-canonical evaluation induced by the set of the least fixed points of Definition 5.2.3, for every $\mathscr{L}_T^{\rightarrow}$-sentence. By Lemma 5.5.2 and the definition of coherence, it is immediate to see that all the results holding for the canonical evaluation $\mathscr{C}$ (Theorems 5.3.10 and 5.3.11, and Corollaries 5.3.12 and 5.3.13) hold for every quasi-canonical evaluation induced by a coherent and exhaustive set of sets of fixed points of Definition 5.2.3 as well. This shows that quasi-canonical evaluations are generalizations of $\mathscr{C}$, as they behave exactly as $\mathscr{C}$ does, with the only exception of assigning a value to possibly more $\mathscr{L}_T^{\rightarrow}$-sentences.

Definition 5.5.4. (Canonical logical consequence)

Let $\mathcal{Q}$ be a set of coherent and exhaustive sets of sets of fixed points of Definition 5.2.3. Define the *canonical logical consequence relation generated by* $\mathcal{Q}$, in symbols $\models^{\mathcal{Q}}_{\mathscr{C}}$, as follows:

$\Gamma \models^{Q}_{\mathscr{C}} \psi$ if and only if for every quasi-canonical evaluation $\mathscr{C}_{\mathcal{G}}$ induced by a

$\mathcal{G} \in Q$, if for every $\varphi \in \Gamma, \mathscr{C}_{\mathcal{G}}(\varphi) = 1$, then $\mathscr{C}_{\mathcal{G}}(\psi) = 1$. ▲

Every canonical consequence relation $\models^{Q}_{\mathscr{C}}$ can be used to describe the introduction and elimination rules of connectives and quantifiers of $\mathscr{L}_T^{\rightarrow}$. First, however, an auxiliary notion is needed to find a genuine condition for conditional introduction.

Definition 5.5.5. (Inverse-determinateness)
For $\varphi \in \mathscr{L}_T^{\rightarrow}$, put: $\mathcal{I}(\varphi) := (\neg\varphi \rightarrow \varphi)$ and call $\mathcal{I}$ an *inverse-determinateness operator*. Following Definition 5.4.1, let $\mathcal{I}\mathcal{I}(\varphi)$ be $\mathcal{I}(\mathcal{I}(\varphi))$, and let $\mathcal{I}^{\underline{n}}(\varphi)$ be a string of $n+1$ iterations of $\mathcal{I}$ applied to φ. Let Nt be the ordinal notation system used in Definition 5.4.1, whose range is O_{Nt}, and define the corresponding hierarchy *à la* Field for O_{Nt}. As for ω iterations of $\mathcal{I}$, let q be the primitive recursive function from positive integers and sentences to sentences s.t. $q(n, \ulcorner\varphi\urcorner) := \mathcal{I}^{\underline{n}}(\varphi)$, and define $\mathcal{I}^{\underline{\omega}}(\varphi)$ as $\exists n T \ulcorner \dot{q}(n, \ulcorner\varphi\urcorner)\urcorner$. ▲

Theorem 5.5.6. (Partial, Bivalent, Limit, Idempotent Inverse-Determinateness)[46]
Let Q be a set of coherent and exhaustive sets of sets of fixed points of Definition 5.2.3. For every $\mathscr{C}_{\mathcal{G}}$, for $\mathcal{G} \in Q$, and every $\varphi \in \mathscr{L}_T^{\rightarrow}$ s.t. $\mathscr{C}_{\mathcal{G}}(\varphi) = k$, for $k \in V$, for each numerical Łukasiewicz value space V (see Subsection 5.2.1), there is a unique ordinal $\delta^V \in O_{Nt}$ (identical to the ordinal δ^V of Theorem 5.4.2) s.t. for all $\delta \in O_{Nt}$ s.t. $\delta \geq \delta^V$:

$$\mathscr{C}_{\mathcal{G}}(\mathcal{I}^{\underline{\delta}}(\varphi)) = \begin{cases} 1 & \textit{if and only if } \mathscr{C}_{\mathcal{G}}(\varphi) \in V \textit{ and } \mathscr{C}_{\mathcal{G}}(\varphi) \neq 0 \\ 0 & \textit{if and only if } \mathscr{C}_{\mathcal{G}}(\varphi) = 0. \end{cases} \tag{5.6}$$

Theorem 5.5.7. (Genuine conditional introduction for canonical consequences)
Let $\mathcal{Z}(\varphi, \psi)$ abbreviate $((\varphi \vee \neg\varphi) \vee \psi) \vee [(\mathcal{I}^{\underline{\delta^V}}(\varphi) \wedge \neg\mathcal{D}^{\underline{\delta^V}}(\psi) \wedge (\varphi \wedge \psi \leftrightarrow \varphi)]$.

[46] The proof is dual to the proof of Theorem 5.4.2.

For every canonical logical consequence relation $\models^{Q}_{\mathscr{C}}$ as in Definition 5.5.4, every $\varphi, \psi \in \mathscr{L}_T^{\rightarrow}$, and every $\Gamma \subseteq SENT_{\mathscr{L}_T^{\rightarrow}}$:

$$\Gamma \models^{Q}_{\mathscr{C}} \varphi \rightarrow \psi \quad \textit{if and only if} \quad \begin{cases} \Gamma, \varphi \models^{Q}_{\mathscr{C}} \psi, \textit{ and} \\ \Gamma \models^{Q}_{\mathscr{C}} \mathcal{Z}(\varphi, \psi). \end{cases}$$ [47]

We can now state introduction and elimination rules for connectives and quantifiers, including the conditional, plus other fundamental inferences.

Theorem 5.5.8. (Introduction/elimination rules for canonical consequences)[48] *For every canonical logical consequence $\models^{Q}_{\mathscr{C}}$ as in Definition 5.5.4, for all sentences $\varphi, \psi, \chi \in \mathscr{L}_T^{\rightarrow}$, we have that:*

- *If $\varphi \models^{Q}_{\mathscr{C}} \psi$ then $\neg\psi \models^{Q}_{\mathscr{C}} \neg\varphi$.*
- *$\neg\neg\varphi \models^{Q}_{\mathscr{C}} \varphi$ and $\varphi \models^{Q}_{\mathscr{C}} \neg\neg\varphi$.*
- *$\varphi, \psi \models^{Q}_{\mathscr{C}} \varphi \wedge \psi$.*
- *$\varphi \wedge \psi \models^{Q}_{\mathscr{C}} \varphi$ and $\varphi \wedge \psi \models^{Q}_{\mathscr{C}} \psi$.*
- *$\varphi \models^{Q}_{\mathscr{C}} \varphi \vee \psi$ and $\psi \models^{Q}_{\mathscr{C}} \varphi \vee \psi$.*
- *If $\varphi \models^{Q}_{\mathscr{C}} \chi$ and $\psi \models^{Q}_{\mathscr{C}} \chi$, then $\varphi \vee \psi \models^{Q}_{\mathscr{C}} \chi$.*
- *If $\varphi \models^{Q}_{\mathscr{C}} \psi$ and $\models^{Q}_{\mathscr{C}} \mathcal{Z}(\varphi, \psi)$, then $\models^{Q}_{\mathscr{C}} \varphi \rightarrow \psi$.*
- *$\varphi, \varphi \rightarrow \psi \models^{Q}_{\mathscr{C}} \psi$.*
- *$\forall x \varphi(x) \models^{Q}_{\mathscr{C}} \varphi(t)$ for every closed $\mathscr{L}_T^{\rightarrow}$-term t.*
- *If $\models^{Q}_{\mathscr{C}} \varphi(t)$ for every every closed $\mathscr{L}_T^{\rightarrow}$-term t, then $\models^{Q}_{\mathscr{C}} \forall x \varphi(x)$.*

[47] The proof of this result and the proof that $\mathcal{Z}(\varphi, \psi)$ yields a genuine condition for conditional introduction follow the structure of the proof of Theorem 4.4.5 (Subsection 4.4.3). The subschema $(\varphi \vee \neg\varphi) \vee \psi$ of $\mathcal{Z}(\varphi, \psi)$ covers the case where φ or ψ have a classical value. The condition $(\mathcal{I}^{\delta^V}_{-}(\varphi) \wedge \neg\mathcal{D}^{\delta^V}_{-}(\psi) \wedge (\varphi \wedge \psi \leftrightarrow \varphi))$, instead, corresponds to both φ and ψ having a non-classical value, with the former different from 0, the latter different from 1, and the former less than or equal to the latter. The whole schema $\mathcal{Z}(\varphi, \psi)$, thus, matches *almost* entirely the conditions for a conditional to get value 1 in a quasi-canonical evaluation (see Theorem 5.3.10), without reducing to them (since, e.g., $\models^{Q}_{\mathscr{C}} \mathcal{Z}(0 = 0, \lambda)$ but $\not\models^{Q}_{\mathscr{C}} 0 = 0 \rightarrow \lambda$).

[48] The proof of this result is an application of Theorem 5.3.10, and I will omit it for space reasons.

- *De Morgan dualities and the classical structural rules hold for* $\models^{Q}_{\mathscr{C}}$.

Moreover, for every $\varphi, \psi, \chi \in \mathscr{L}_T^{\rightarrow}$, *if* ψ *and* χ *are alike except that* ψ *has an occurrence of* φ *of* $\mathscr{L}_T$ *where* χ *has an occurrence of* $T\ulcorner\varphi\urcorner$, *then:*

$$\psi \models^{Q}_{\mathscr{C}} \chi \quad \textit{and} \quad \chi \models^{Q}_{\mathscr{C}} \psi.$$

5.6 Summary of the chapter

In this chapter I presented an evaluation, the *canonical evaluation*, devised to address FIELD'S PROGRAM and to provide a novel and arguably satisfactory account of truth-theoretic paradoxes, avoiding some of the difficulties of Kripke's theory, of Field's theory, and of the ŁK-construction.

In Section 5.1, I introduced my graph-theoretic approach, via some heuristics (Subsection 5.1.1), a formal presentation (Subsection 5.1.2), and examples (Subsection 5.1.3). In Section 5.2 I defined the evaluations of nodes in semantic graphs: after specifying the reading I adopted for connectives and quantifiers (Subsection 5.2.1), I introduced the evaluations of nodes informally (Subsection 5.2.2), and then formally (Subsection 5.2.3). In Section 5.3, I analyzed some structural properties of semantic graphs (Subsection 5.3.1), using them to define the canonical evaluation and to prove its properties (Subsection 5.3.2). Section 5.4 explored the account of determinateness given by the canonical evaluation, and Section 5.5 associated a notion of logical consequences with the present construction, providing a genuine rule of conditional introduction that generalizes the rule given by the ŁK-construction.

Chapter 6

The Canonical Evaluation and its Philosophical Import

6.1 Introduction

In this chapter, I will argue that the canonical evaluation provides an appropriate semantics for languages such as $\mathscr{L}_T^{\rightarrow}$ and a suitable handling of the truth-theoretic paradoxes affecting them. More in particular, I will argue that the canonical evaluation and its quasi-canonical variants address FIELD'S PROGRAM in a satisfactory way and I will contend that, thanks to its peculiarities (i.e. assigning values in Ł or no value at all to some sentences), the canonical evaluation presents an interesting and acceptable way to account for and treat truth-theoretic paradoxes.

6.2 The canonical evaluation and FIELD'S PROGRAM

I will now argue that the canonical evaluation is a satisfactory candidate to address FIELD'S PROGRAM. Recall that FIELD'S PROGRAM is the search for models that preserve Kripke's theory and add to it a *well-behaved conditional* connective.

6.2.1 The canonical evaluation preserves Kripke's theory

Lemma 6.2.1. (Inclusion of the least Kripke fixed point in $\mathscr{C}$)[1]

For every $\varphi \in \mathscr{L}_T^{\rightarrow}$,

$$\varphi \in \mathfrak{E}_\Phi \text{ if and only if } \mathscr{K}_{\emptyset,\emptyset}(\varphi) = 1 \text{ only if } \mathscr{C}(\varphi) = 1,$$

$$\varphi \in \mathfrak{A}_\Phi \text{ if and only if } \mathscr{K}_{\emptyset,\emptyset}(\varphi) = 0 \text{ only if } \mathscr{C}(\varphi) = 0.$$

The least Kripke fixed point is included in every other Kripke fixed point, and the same holds for the respective Kripke evaluations. So, it seems fair to say that the value assignments given by least Kripke fixed point embody the most fundamental aspect of Kripke's construction, and the canonical interpretation recovers it fully.

The algebraic meaning of connectives and quantifiers we find in Kripke's theory is recovered by the canonical evaluation, as the evaluation clauses of Kripke's theory for $\neg$, $\wedge$, $\vee$, and $\forall$ are preserved by $\mathscr{C}$.[2] Moreover, the introduction and elimination rules of $\neg$, $\wedge$, $\vee$, and $\forall$ that result for the notions of logical consequence I explored for Kripke's theory are recovered in the canonical evaluation.[3] Finally, the truth predicate, as interpreted by $\mathscr{C}$, respects naïveté as in Kripke's theory, since both theories respect weak naïveté and intersubstitutivity of truth.[4]

6.2.2 All the connectives and quantifiers are well-behaved

In Section 2.7, I gave some criteria for the good behavior of connectives and quantifiers, called **(W1)**, **(W2)**, and **(W3)**:[5] the canonical evaluation makes the connec-

[1]The proof is an easy induction, using Theorem 5.3.11 and the construction of $\mathfrak{I}_\Phi$.

[2]Compare Lemma 2.4.3 and Theorem 5.3.11.

[3]See Definition 2.5.7 and Remark 2.5.8, and compare Subsection 2.5.2 and Theorem 5.5.8.

[4]Theorem 2.4.1, Corollary 2.4.2 for Kripke's theory, Theorem 5.3.11, Corollary 5.3.12 for $\mathscr{C}$.

[5] **(W1)** requires that every connective and quantifier has a clear and simple *meaning*. E.g., For example, it should be possible to describe the meaning of every connective and quantifier via some simple algebraic rule (given by the semantics), or to axiomatize adequately the $\mathscr{L}_T^{\rightarrow}$-sentences validated by the semantics, or to characterize every connective and quantifier via suitable introduction and elimination rules (validated by the semantics).

(W2) requires that every connective and quantifier, as interpreted by the semantics, should be closed under suitable introduction and elimination rules, that approximate the classical ones.

tives and quantifiers of $\mathscr{L}_T^{\rightarrow}$ well-behaved according to them. For brevity, I will focus on $\rightarrow$, but the following points apply to $\neg$, $\wedge$, $\vee$, and $\forall$ as well.

1. As to **(W1)**, the canonical evaluation confers a clear and simple meaning to the conditional: it is Łukasiewicz' conditional, rendered partial via the strong Kleene partiality paradigm.[6] The Łukasiewicz reading of $\rightarrow$ is recovered for every conditional sentence that receives a value (in the case of equational values, the Łukasiewicz meaning is visible in the very equations assigned to conditional sentences). So, the canonical evaluation generalizes the reading of the conditional as a *tool to compare semantic values* (restricted to the sentences that receive a value) to every Łukasiewicz value space.[7]

2. As to **(W2)**, the classical introduction and elimination rules for $\rightarrow$ are approximated as closely as possible by the canonical logical consequence,[8] improving on the conditional introduction of the ŁK-construction. Also, the conditional introduction rule allows us to express genuinely the difference between conditional assertion and assertion of the conditional, via an object-linguistic schema that generalizes the rule of the ŁK-construction to every Łukasiewicz numerical value space V.[9]

3. Also **(W3)** is satisfied by the canonical evaluation. In fact, the Łukasiewicz reading of connectives and quantifiers is recovered, together with naïveté, *as much as possible* by $\mathscr{C}$, and so are the pre-formal intuitions that go with these interpretations (e.g. the order-theoretic conception of connectives and quantifiers). The construction I employed to define the canonical evaluation,

(W3) requires that conferring a meaning on the connectives and quantifiers, as specified in **(W1)**, should allow us to render (as much as possible) the intuitions we aimed at capturing with the development of our semantic theory.

[6] See Subsection 4.2.3.

[7] This reading was discussed in Subsections 3.5.1 and 4.4.2.

[8] See Definition 5.5.4 and Theorem 5.5.8.

[9] Compare Theorem 4.4.5 and subsequent remarks with Theorem 5.5.7 and footnote 5.5.7.

however, is not in any way specific to Łukasiewicz semantics. One could modify Definition 5.2.3 using different algebraic clauses for connectives and quantifiers: since Theorem 5.3.4 would still hold, the evaluations defined on nodes resulting from the modified construction would still be uniform (in the sense of Theorem 5.3.7), and a correspondingly modified version of the canonical evaluation would obtain. Any such modified version of the canonical evaluation would preserve *as much as possible* the chosen algebraic reading of connectives and quantifiers, and thus every pre-theoretic intuition codified in it, together with naïveté.[10] The construction that leads to the canonical evaluation, so, is more flexible, easily modifiable, and applicable than Kripke's theory, Field's theory, or the ŁK-construction.

In addition to addressing FIELD'S PROGRAM, the canonical evaluation realizes some intuitions about languages such as $\mathscr{L}_T^{\rightarrow}$ that are, I argue, important to provide a satisfactory understanding of naïve truth and informative solutions to paradoxes. I will now explore these intuitions and their philosophical relevance.

6.3 Łukasiewicz equations as semantic values

One of the peculiar aspects of $\mathscr{C}$ is the use of Łukasiewicz equations as semantic values. A sentence is assigned a Łukasiewicz equation when its semantic graph generates a fully determinate equation system that either has no solution in the chosen numerical value space V, or has more than one solution in it.

6.3.1 No solutions

If a sentence is assigned an equational value because its equation system has no solutions in V, it means that there is just no way of assigning to that sentence

[10] Some other pre-theoretic intuitions embodied in this construction (e.g. about Yablo-like cases) are independent from the possibility to adopt other evaluation schemes for numerical assignments.

a numerical value that agrees with $\mathbb{N}$ on atomic arithmetical sentences and that respects weak naïveté and the Łukasiewicz reading of connectives and quantifiers.

We know that there is no way of assigning a value in V to every sentence, respecting weak naïveté and Łukasiewicz semantics.[11] Sentences that generate unsolvable equation systems could be seen as residuals generated by the choice of adding the Łukasiewicz conditional to the stock of our connectives. While every sentence of the conditional-free language $\mathscr{L}_T$ can receive a value consistently with weak naïveté already when $V = \{0, 1/2, 1\}$ (as Kripke's theory shows), the Łukasiewicz conditional is so strong that not even taking $V = [0,1]$ suffices to give a value to every sentence. Sentences that are assigned equational values because their equation systems have no solutions are nonetheless informative: the canonical evaluation *treats them explicitly* and thus *exhibits* the effects of the addition of naïve truth to the language $\mathscr{L}_T^{\rightarrow}$ interpreted by Łukasiewicz semantics.

6.3.2 Too many solutions

The canonical evaluation assigns values in Ł also to sentences whose evaluation generates fully determinate equation systems that have many solutions. This seems an informative picture: for such sentences, Łukasiewicz equations are *as close as we can possibly get* to a numerical value. The semantics displays all the relevant information that can be used to give a numerical value to sentences of this kind: it is simply that the chosen interpretation of connectives, quantifiers, and the truth predicate does not single out a unique value for them. So, the best we can do is to state all their relations with other sentences.

A noteworthy case of this kind is the truth-teller (but the following points apply to all the relevantly similar cases). $\mathscr{C}$ assigns τ its Łukasiewicz equation, i.e. $\mathbf{u}_\tau =_{\mathbf{U}} \mathbf{u}_\tau$, and this seems the most informative value one can give to it, since it can

[11] See Theorems 4.1.2 and 5.4.3.

be used to explain why τ can be consistently given any numerical value.

Kripke's theory, by contrast, treats the truth-teller variably: τ has no value in the least Kripke fixed point and has value 0 or 1 in some consistent non-minimal fixed point.[12] Neither of these solutions is acceptable. While I accept that some sentences should be assigned no value, this is not the case for τ, whose relational structure is so simple that it could and it should be brought to light (Figure 5.3), being identifiable by very weak standards. On the other hand, the fact that it is consistent to assign to τ any numerical value available, including 0 or 1, suggests that we should not give it a numerical value: as argued in Subsection 2.5.1, such extreme fluctuation should not be allowed if numerical values, especially 0 and 1, are to indicate something (e.g. acceptance and rejection).

One can reply that Kripke's theory characterizes the behavior of τ via the collection of Kripke fixed points, but I tried to show in Subsection 2.5.1 that this characterization is unsatisfactory: Kripke fixed points are the interpretations for $\mathscr{L}_T^{\rightarrow}$-sentences produced by Kripke's theory, and no such interpretation can account for the main semantic trait of the truth-teller, i.e. that it can consistently take any numerical value.[13] The canonical evaluation, instead, does exactly this, by assigning τ an equation that has every numerical value in V as a solution.

The treatment of τ in Field's theory derives from that of Kripke's theory: $\mathscr{F}(\tau) = \mathbf{1/2}$, but there are variants of Field's theory, obtained via non-minimal Kripke fixed points at revision stages, where τ has value $\mathbf{1}$ or $\mathbf{0}$. So, as argued in Section 3.7, the treatment of the truth-teller is as inadequate in Field's theory as it is in Kripke's theory, and the canonical evaluation improves on both accounts.

Another aspect of τ is accounted for by $\mathscr{C}$. τ is often glossed informally as "the sentence that says of itself that it is true", or the like. This paraphrase is sloppy, but in view of the definition of τ, one might want to say that τ is somehow

[12] See Subsection 2.5.1. The ŁK-construction, moreover, handles τ in an analogous way.

[13] A similar point applies to Kripke $\mathscr{A}$-consequences (Definition 2.5.7).

characterized by its *equivalence* with $T\ulcorner\tau\urcorner$, for a suitable notion of equivalence.

This characterization, however, is not easily obtained in a *semantic* theory of naïve truth. Suppose one characterizes φ and ψ as equivalent whenever they have the same value: then, every theory accepting weak naïveté declares φ equivalent to $T\ulcorner\varphi\urcorner$, for every sentence φ (possibly *modulo* partiality). Clearly we cannot capture the desired feature of τ in this way. A stronger notion of equivalence will not do either. For suppose that we declare two sentences φ and ψ equivalent if and only if $\varphi \leftrightarrow^\star \psi$ gets the/a designated value, for some biconditional $\leftrightarrow^\star$. Virtually every theory of truth validates $0=0 \leftrightarrow^\star T\ulcorner 0=0\urcorner$, but we do not want to say that $0=0$ is characterized by its equivalence with $T\ulcorner 0=0\urcorner$ in the same way in which τ is characterized by its being somehow equivalent to $T\ulcorner\tau\urcorner$.

One might resort to other considerations to account for the targeted feature of τ: e.g., unlike $0=0 \leftrightarrow^\star T\ulcorner 0=0\urcorner$, the equivalence $\tau \leftrightarrow^\star T\ulcorner\tau\urcorner$ is provable in several axiomatic theories without using any axiom or rule for truth, i.e. simply via the diagonal lemma. Such considerations, however, bring proof-theory or syntax into the picture. My point is that a *semantic* theory should allow us to capture the targeted feature of τ, since it is plausible to argue that τ can consistently be assigned any semantic value *because* it only consists of the application of T to a term denoting the code of τ, and so the targeted equivalence between τ and $T\ulcorner\tau\urcorner$ is fundamental to account for the semantics of τ.

The canonical evaluation solves the problem straightforwardly: τ receives a special semantic value that yields the desired characterization. A sentence φ receives the value $\mathbf{u}_\varphi =_\mathbf{U} \mathbf{u}_\varphi$ if and only if φ has the form Tt and $t = \ulcorner\varphi\urcorner$. The "equivalence" between τ and $T\ulcorner\tau\urcorner$ is codified in the semantic value $\mathbf{u}_\tau =_\mathbf{U} \mathbf{u}_\tau$. This kind of solution applies to all truth-teller-like cases.

Some objections to my treatment of truth-teller-like cases arise immediately:

(o1) The canonical evaluation captures the fact that τ can have any numerical value, but it does not capture the fact that λ cannot have any classical value.

(o2) If we want to say that τ is characterized by its being somehow equivalent to $T\ulcorner\tau\urcorner$, we should also want to say that, e.g., λ is characterized by its being somehow equivalent to $\neg T\ulcorner\lambda\urcorner$, but $\mathscr{C}$ makes the two cases disanalogous.

Both these objections, however, can be easily addressed:

(r1) λ is assigned an equation whose *unique* solution is a non-classical value. This seems to answer (o1), but we can capture the desired feature of λ even more explicitly: setting $V = \{0,1\}$, λ is given an equation with no solutions.

(r2) Setting $V = \{0,1\}$, λ receives an equation that provides the desired characterization, as in the case of τ. If V is as specified in Subsection 5.2.1, instead, then the cases of λ and τ *are* disanalogous. The structural features of τ are semantically relevant because they make τ compatible with any numerical value assignment. In the case of λ we have an opposite situation, as the structure of λ and our interpretation of $\neg$ and T single out a unique value.[14]

Summing up, via the canonical evaluation (and its quasi-canonical variants) one can distinguish sharply between truth-teller-like cases and arguably different cases such as the liar. Truth-teller-like cases receive special semantic values that avoid the difficulties related to numerical values and that express their semantic behavior and their main structural peculiarities.[15]

[14] To emphasize the disanalogy, let κ^g be $Tt_{\kappa^g} \to 0 = 0$, where $t_{\kappa^g} = \ulcorner Tt_{\kappa^g} \to 0 = 0\urcorner$ (I owe this example to Vann McGee). There is a sense in which κ^g is somehow equivalent to the sentence that says that, if κ^g is true, then $0 = 0$. However, if we believe that conditionals with 1-valued consequents have value 1 (as in the strong Kleene paradigm), we believe that $Tt_{\kappa^g} \to 0 = 0$ has value 1; so, in this case, the above "equivalence" is not relevant from the semantic standpoint.

[15] One could refine $\mathscr{C}$ further by altering the equational values. For example, one could distinguish the "no solutions" and the "too many solutions" cases by using, as semantic values, *the set of solutions of equation systems*, not single equations (this suggestion was made by Hartry Field). A cost of this option is that the value space would become less intuitive. A related idea is followed in [Pailos and Rosenblatt 2015], albeit in the setting of poly-valued logics: the theory is technically interesting, but the conditional is difficult to interpret.

6.4 Circularity and semantic value assignments

6.4.1 Circularity and self-reference

The canonical evaluation employs a notion of circularity in its use of loops, and a notion of circularity is sometimes claimed to be closely related to "self-reference" or even to "paradoxicality". This is a specialized issue that would require a thesis on its own, so I will only make a few remarks related to the canonical evaluation.

Much work on circularity and self-reference in truth-theoretic paradoxes deals with Yablo's paradox and its variants.[16] Here is the semantic graph of the 0-th Yablo sentence: it will be a guide to how $\mathscr{C}$ treats every Yablo sentence.

$$\begin{tikzcd}
& & \forall k > 0 \neg T \ulcorner \upsilon_k \urcorner \arrow[dl] \arrow[drr] & & & \\
& \neg T \ulcorner \upsilon_1 \urcorner \arrow[d] & \cdots & & \neg T \ulcorner \upsilon_j \urcorner \arrow[d] & \cdots \\
& T \ulcorner \upsilon_1 \urcorner \arrow[d] & & & T \ulcorner \upsilon_j \urcorner \arrow[d] & \\
& \forall k > 1 \neg T \ulcorner \upsilon_k \urcorner \arrow[dl] \arrow[dr] & \cdots & & \forall k > j \neg T \ulcorner \upsilon_k \urcorner \arrow[dl] \arrow[dr] & \cdots \\
\vdots & & \vdots & \vdots & & \vdots
\end{tikzcd}$$

None of the semantic graphs of any Yablo sentence features a loop, i.e. a *circularity* in my sense. The canonical evaluation, however, gives also a sense in which Yablo-like sentences are not *self-referential*, as I will now argue.

Let the *components* of a sentence φ be the sentences obtained considering iteratively the proper subsentences of φ and the results of truth-theoretic disquotation.[17] In terms of semantic graphs, we can identify the components of a sentence φ as

[16] See [Yablo 1993], [Priest 1997], [Beall 2001], [Leitgeb 2002], [Bueno and Colyvan 2003], [Ketland 2004], [Cook 2004], [Cook 2006], [Schlenker 2007], [Cook 2014].

[17] See also also [Hansen 2015], p. 58.

the sentences labelling those nodes of $\mathtt{N}_\varphi$ that receive at least one arrow. Of course, even the root itself may receive some arrows.[18]

Consider now the definition of semantic graphs (Definition 5.1.2), and in particular the clause $(\mathtt{Loop})$. Using the notion of component, we can identify a precise sense in which $(\mathtt{Loop})$ checks for self-reference. $(\mathtt{Loop})$ applies when, in a semantic graph, an occurrence of σ would generate another occurrence of σ as successor, if not blocked: such σ may be considered self-referential, since this situation occurs exactly when at least one of the nodes reachable from the first occurrence is labelled with Tt_σ for a term t_σ s.t. $t_\sigma = \ulcorner\sigma\urcorner$. As it is customary to see Gödel coding as modeling the syntactic device of quotation, to an extent at least, it seems acceptable to say that Tt_σ *mentions* σ.[19]

This conclusion is reinforced by the fact that, in formal theories of truth, we use T to represent the *predicate* "… is true", that applies to names of sentences, and the Gödel coding provides the naming device. So, Tt_σ could be semi-formally paraphrased as "«σ» is true". This reading is represented writing Tt_σ as $T\ulcorner\sigma\urcorner$. This notation, albeit not incorrect, is slightly misleading, as it suggests that σ is syntactically a sub-expression of $T\ulcorner\sigma\urcorner$, which is not. Nevertheless, the adoption of basically any form of naïveté commits us to consider σ a component of Tt_σ.

Thus, that one of the components of σ mentions σ may be understood as the fact that σ *includes a reference to itself*. This seems enough to justify the claim that σ, amongst the other things it possibly says, *refers to* σ, i.e. *it refers to itself*, i.e. *it is self-referential*. We can put forward the following definition:

$(\mathtt{SR})$ A sentence φ is *self-referential* iff φ belongs to the components of φ.

[18] The notion of components of a sentence φ is not sensible to the choice of the root of $\langle \mathtt{N}_\varphi, \mathtt{S}_\varphi \rangle$. By Theorem 5.3.4, for every semantic graph $\langle \mathtt{N}_\psi, \mathtt{S}_\psi \rangle$ including a node $\mathtt{v}$ labelled as φ, the set of sentences labelling the nodes receiving at least one arrow in $\langle \mathtt{N}_\varphi, \mathtt{S}_\varphi \rangle$ is *identical* to the set of sentences labelling the nodes receiving at least one arrow in a path starting from $\mathtt{v}$ in $\langle \mathtt{N}_\psi, \mathtt{S}_\psi \rangle$. So, we can speak interchangeably of a node labelled with a sentence and of that sentence.

[19] For an in-depth analysis of self-reference, also in relation to the chosen coding, see [Halbach and Visser 2014a] and [Halbach and Visser 2014b].

Due to our naïve interpretation of truth, such self-reference is explicitly brought to light, as all the instances of truth predicate are disquotationally eliminated: if the semantic graph of σ includes the sentence Tt_σ as above, this sentence is then further analyzed into σ, as indicated by the looping arrow that this generates. But then one can claim that *all and only the self-referential sentences are identified via the clause* (Loop). If σ is self-referential according to (SR), it belongs to a loop generated by the clause (Loop) together with a sentence Tt_σ for a term t_σ s.t. $t_\sigma = \ulcorner \sigma \urcorner$, and *vice versa*. Under this interpretation, no Yablo-like sentence comes out as self-referential, and the clause (Loop) and its interpretation provide a clear and well-motivated explanation of this result.

Yablo sentences count as circular or self-referential in other accounts. [Priest 1997], for example, claims that any sentence that results from a fixed-point construction is circular and self-referential. Priest then analyzes Yablo sentence and shows that they are all fixed points of a certain formula, concluding that they display the same kind of circularity/self-reference that is given in the liar sentence. [Cook 2006] refines Priest's claim, by distinguishing between various kinds of circularity/self-referentiality that arise in conjunction with different notions of fixed point used in mathematics and logics.[20]

Priest's and Cook's analyses do justice to the intuition that all Yablo sentences "say the same thing". Identifying self-referentiality with a certain fixed-point property, however, they give a non-semantic account of this notion.[21] (SR), by contrast, provides a semantic characterization of self-referentiality, since it revolves around the semantic relations of *mentioning* or *referring to*: as such, this notion of self-reference is something that a semantic theory should address.

[20][Cook 2014] elaborates on [Cook 2006], retaining the conclusion that Yablo's paradox (in its original version) is circular, but circularity alone does not account for paradoxicality.

[21]This is often acknowledged by the proponents of the fixed-point analysis of self-referentiality: Cook, for example, argues that the right notion of fixed point to use to analyze self-referentiality comes in terms of provability in a suitable formal theory (see [Cook 2006], p. 122-124).

Interestingly, a Priestian analysis of the self-referentiality of Yablo's paradox may be based also on the *semi-formal* presentations of the paradox given by Yablo. Consider the following passage, the *locus classicus* for Yablo's paradox.[22]

> Imagine an infinite sequence of sentences $[\sigma_1]$, $[\sigma_2]$, $[\sigma_3]$, ..., *each to the effect that every subsequent sentence is untrue*:
>
> (σ_1) for all $k > 1$, $[\sigma_k]$ is untrue
>
> (σ_2) for all $k > 2$, $[\sigma_k]$ is untrue
>
> (σ_3) for all $k > 3$, $[\sigma_k]$ is untrue
>
> $\vdots$
>
> ([Yablo 1993] p. 251, the emphasis is mine)

In Yablo's semi-formal presentation, we are given a description of the syntactic structure of *every* sentence in the infinite list (in the passage I italicized). So, one might adopt Priest's reasoning and conclude that Yablo sentences are self-referential also on the basis of Yablo's semi-formal presentations, as they include a description of the common structural feature of every Yablo sentence. Also here, we are given *syntactic* information that one might use to claim, following Priest, that every Yablo sentence says the same thing, or the like.

Now, one may of course accept the idea that, in order to identify self-reference, one needs to resort to extra-semantic notions, as Priest and Cook do. In this case, however, it is unclear why a semantic theory should address self-reference, or use it in assigning semantic values. The apparatus of semantic graphs, instead, provides a treatment of circularity and self-reference that has a clear semantic rationale. The resulting account may be unsophisticated, but it has some virtues: the difficulties of the syntactic characterizations of self-reference are avoided, and the informal ideas of circularity and self-reference are absorbed in the fairly well-understood array

[22] But see also the presentations in [Yablo 1985], p. 340, and [Yablo 2006], p. 140.

of notions of well-foundedness of graphs,[23] given their identification with loops, conferring them a structurally precise guise.

6.4.2 The roots of semantic values assignments

If a sentence is given no value by $\mathscr{C}$, we have seen, its semantic graph has at least one infinitely descending path. How does one motivate the idea that whether a sentence is assigned a value should revolve around infinitely descending paths? To answer this question, I will outline and defend a general view on the assignment of semantic values, from which the valuelessness of Yablo-like cases follows.

To introduce my view, let me repeat that my main purpose in studying languages such as $\mathscr{L}_T^{\rightarrow}$ is to understand the semantics of simple fragments of natural languages (featuring mathematical and semantic vocabulary) and the paradoxes that affect them. The paradigm in which I worked so far is a broadly compositional view according to which, in order to assign a value to φ, we should consider the values assigned to its fundamental elements. While it is generally agreed that the fundamental elements of a sentence are its sub-sentences, our acceptance of naïveté commits us to consider also the results of truth-theoretic disquotation; in a word, the fundamental elements of a sentence are its components.[24]

Semantic graphs and the evaluations of nodes are a way to assign a value to φ based on the values assigned to its components. The graph of φ, in fact, yields the sub-sentences of φ and the results of truth-theoretic disquotations, then the sub-sentences of the resulting sentences and the new results of disquotation, and so on. In this process, several sentences are found, together with the relations holding between them (including loops). *At some point*, it may happen that information emerges that is sufficient to assign a value to some components of φ, and possibly

[23]See the idea of groundedness* (Definition 5.1.5).

[24]A sentence of $\mathscr{L}_T^{\rightarrow}$ has no components if it is an atomic arithmetical sentence, or it has the form Tt for a term t that does denote the code of a sentence: in any such case, however, we have already sufficient information to assign a semantic value to it.

to φ as well – this task is performed by the evaluation of nodes.

Semantic graphs and the evaluations of nodes are tools to implement the compositional view that I want to advocate, and that can be formulated as follows:

(IAC) For every sentence φ, the assignment of semantic value to φ is determined by the semantic values of the components of φ and by the semantic relations between them, that are (respectively) assigned and discovered *at some point* in understanding and unravelling the structure of φ.[25]

A difference between (IAC) and the usual compositionality principles is that the former does not concern the *meaning of expressions*, but the *semantic value of sentences*. Consider the following classical formulation of compositionality:

(COMP) The meaning of a complex expression is determined by the meanings of its constituents and by its structure. ([Szabò 2000], p. 3)

More specific kinds of compositionality, such as (IAC), are sometimes seen as special cases of (COMP). Compositional principles about semantic values are sometimes called *compositionality of reference*, in a somewhat Fregean style.[26] It is clear, however, that in the present work a semantic view has to be articulated around semantic values of sentences, since they are amongst the main object of investigations for semantic theories of truth, unlike the meanings of expressions. Moreover, following the approach I have adopted in the course of this work, (IAC) is concerned with the *assignment* of semantic values to sentences, which is one of the central preoccupations of semantic theories of truth: everyone agrees on how to assign a semantic value to arithmetical sentences (in the case of $\mathscr{L}_T^{\rightarrow}$), the real source

[25]I do not endorse any specific view on the nature of the *determination* mentioned in (IAC): these matters are extraneous to the purely semantic compositional principle I endorse, so I remain agnostic about them. One is free, for example, to understand the *being determined* in purely operational terms: the semantic values assigned to a complex expression are the result of some algebraic manipulation of the semantic values assigned to its components.

[26]See [Szabò 2000], pp. 3-9.

of disagreement and the space of theorizing is how to assign values to sentences involving the logical operators and T.

The label "IAC" stands for *idealized assignability compositionality*. The discovery of the components of φ and of the relations holding between them, and the assignment of semantic values should be understood as carried out by an idealized agent, that has infinitary computational powers. For example, an idealized agent can assign value 1 to a universally quantified sentence because she considers all its instances, which are infinitely many, and assigns value 1 to all of them. I assume that this idealization is not out of place, since (IAC) is designed to capture a semantic view. What matters for (IAC) is that, in unravelling and understanding the structure of a sentence, *at some point* an idealized agent can start to assign semantic values and to identify semantic relations between the components of φ, possibly leading to assigning a semantic value to φ itself – it does not matter whether in order to do this our agent has to check infinitely many sentences.

The peculiarity of (IAC) becomes important when the truth predicate is given a naïve reading. The usual picture of compositionality, in fact, does not contemplate cases where *at no point* the components of a sentence and the relations between them are not fully identified. However, a naïve reading of truth entails the existence of loops and of never ending processes, in unraveling the components of a sentence – i.e. infinitely descending paths in semantic graphs. The latter case forces us to include the possibility that *at no point* the components of a sentence and the semantic relations between them are fully identified.

(IAC) excludes from semantic value assignment every Yablo sentence and relevantly similar cases. Moreover, under several evaluation schemes, the adoption of a strong Kleene partiality paradigm rules also out cases such as the conjunction of $0 = 0$ and a Yablo sentence, where an idealized agent can give a semantic value to a component of this sentence at some point, but this does not suffice to evaluate the entire sentence. Dually, in the disjunction of $0 = 0$ and a Yablo sentence an idealized agent finds enough information to give value 1 to the whole sentence.

These ideas are formalized and implemented in the canonical evaluation, but the view outlined in (IAC) is independent from it. Now I will argue that (IAC) provides an acceptable description of how the assignments of semantic values should be carried out, in simple languages such as those targeted in this work. At the same time, I will defend (IAC) against the objection that sentences should be assigned semantic values even if there is no point, in the unravelling and understanding of its structure, at which it is possible to start assigning values to its components and to identify semantic relations between them, as in Yablo-like cases.[27]

Consider the following informal game. There are two players, P (the Producer) and E (the Evaluator): P has to produce sentences, E has to evaluate them. P can produce how many sentences she likes, and the game can be extended indefinitely; of course, she can also stop (e.g., by uttering a conventional sentence that does not count amongst those to be evaluated). In order to evaluate sentences, E does not have to wait for P to stop: for example, if P says "$0 = 0$" and then goes on with other sentences, E can evaluate $0 = 0$ right after P utters it. E wins the game if, for every non-empty subset of the sentences uttered by P (either after P stops producing new sentences or while she is still producing sentences), E evaluates them in such a way that: (I) E's assignments of semantic values are correct according to the semantics she chooses, and (II) E's assignment cannot be rendered incorrect by any sentence P could possibly produce after they are made.[28] As to (I), of course, one should at least assume weak naïveté, to be consistent with the setting we are work in. The constraints (I) and (II) are reasonable enough: E's evaluation must

[27] For example, Gaifman's theory (see [Gaifman 1992]) includes the resources to address explicitly this particular feature of the Yablo sentences, and was designed having (also) such cases in mind (personal communication with Haim Gaifman). On the difficulties to evaluate Yablo-like sentences from a contextualist point of view and from Gaifman's perspective, see [Yi 1999].

[28] Of course, condition (II) is only relevant for those value assignments that E makes while P has not stopped producing sentences.

be coherent and E should only assign semantic values that cannot turn out to be wrong according to the semantics she chooses.

Let's now consider a version of Yablo's paradox in terms of the above game.

> **Yablo's Game.** Here is a description of the first n sentences P utters:
>
> > "Call this very sentence "*Sentence* 1": all the sentences I utter after *Sentence* 1 are not true."
> >
> > "Call this very sentence "*Sentence* 2": all the sentences I utter after *Sentence* 2 are not true."
> >
> > $\vdots$
> >
> > "Call this very sentence "*Sentence n*": all the sentences I utter after *Sentence n* are not true."
>
> The vertical dots stand for sentences structured in a way corresponding to *Sentence* 1, *Sentence* 2, and *Sentence n*.
> After these n sentences, P goes on similarly: she keeps saying sentences structured as *Sentence* 1, *Sentence* 2, and *Sentence n*, *without ever stopping* and *without ever uttering a different kind of sentence*. In fact, P produces larger and larger fragments of a Yablo list.[29]

Even granting that E is an idealized agent as described above, is she ever in a position win **Yablo's Game**? Under the following (very modest) assumptions on the semantics adopted by E, she is not:[30]

(A1) There are at least two values, $\mathfrak{V}$ and $\mathfrak{F}$. $\mathfrak{V}$ is the only *designated* value. At least one sentence receives $\mathfrak{V}$ or $\mathfrak{F}$. No sentence receives many values.

[29] **Yablo's Game** shares some aspects with the version of Yablo's paradox in [Sorensen 1998].

[30] Without loss of generality, I suppose that the following conditions are given in a classical metatheory for the semantics adopted by E.

(A2) A universally quantified sentence has value $\mathfrak{V}$ only if all its instances have value $\mathfrak{V}$.

(A3) A negation has value $\mathfrak{V}$ ($\mathfrak{F}$) if and only if the negand has value $\mathfrak{F}$ ($\mathfrak{V}$).

(A4) Universal and existential quantifiers are interdefinable, i.e. $\forall x\varphi(x)$ has the same semantic value as $\neg\exists x\neg\varphi(x)$; $\exists x\varphi(x)$ has the same value as $\neg\forall x\neg\varphi(x)$.

(A5) An existentially quantified sentence has value $\mathfrak{V}$ if and only if at least one of its instances has value $\mathfrak{V}$.

(A6) Weak naïveté.

Suppose that, *at some point* while P is producing sentences, E assigns a value to one of them. Such sentence, then, will be a *Sentence k*, for some natural number k. By (A1), there are three main cases.

1. Suppose that E assigns value $\mathfrak{V}$ to *Sentence k*. Then, by (A2), (A3), and (A6), E also assigns value $\mathfrak{F}$ to every sentence that P produced/will produce after *Sentence k*, and E's assignment would turn out to be wrong if P uttered something that E's semantics validates (which exists by (A1) and (A3)).

2. Suppose that E assigns value $\mathfrak{F}$ to *Sentence k*. Then, E assigns value $\mathfrak{V}$ to some sentence φ that P utters after *Sentence k*, by (A3), (A4), (A5), and (A6). Since P produces a Yablo sequence, φ is some *Sentence j*, for $k < j$. Since *Sentence j* has value $\mathfrak{V}$, all the sentences uttered by P after *Sentence j* have value $\mathfrak{F}$, by (A2), (A3), and (A6). In particular, thus, *Sentence* $j+1$ (which P utters right after *Sentence j*) has value $\mathfrak{F}$. Then, E must give value $\mathfrak{V}$ to at least one sentence that P utters after *Sentence* $j+1$, by (A4) and (A5), making her assignments incoherent.[31]

[31] The latter part of this reasoning is very much as in [Yablo 1993].

3. Finally, suppose the semantics adopted by E features some extra semantic values and that E gives one such value to *Sentence k*. As in case 1, E's assignment would turn out to be wrong if P uttered a sentence that E's semantics validates (which exists by (A1) and (A3)), as E would then be forced to assign value $\mathfrak{F}$ to *Sentence k* (by (A3), (A4), (A5), and (A6)).

As k is arbitrary, this reasoning applies to every sentence uttered by P and to every non-empty subsets of them. Thus, *modulo* our assumptions (A1)-(A6), E cannot win **Yablo's Game**. In fact, something stronger holds: *at no point E can make an* assignment of semantic values that respects conditions (I) and (II) above.[32]

The crucial feature of **Yablo's Game** is that E is never in a position to determine that P is producing a Yablo sequence. This is a radical difference with Yablo's original presentations, where we are explicitly told how the sequence goes, or with the formalized versions of the paradox, as shown by Priest's analysis.

In **Yablo's Game**, in fact, things are even more complicated. For one might think that the situation of **Yablo's Game** would have been different if, at some point, P had said something along the following lines:

> "Call this very sentence "*Sentence Y*": with the exception of *Sentence Y*, every other sentence I will produce in this game will have the following form "Call this very sentence "*Sentence k*": all the sentences I will utter after *Sentence k* are not true", using growing natural numbers in my enumeration."

[32]One might object that my condition (II) is too restrictive, as it rules out the possibility to change one's assignment of values, e.g. as in revision semantics. This objection is, however, misplaced. E is free to adopt any semantics she likes, including revision semantics. If P utters a liar sentence, an evaluator E who adopts a revision semantics *à la* Gupta-Belnap does have enough information to determine its revision sequence, or a value that represents its revision sequence, and this situation cannot possibly be altered by any other sentence uttered by P. This is in line with what revision theorists do: they revise revision values, not revision sequences. So, condition (II) is not unfavorable to revision theorists. The problem pointed at by **Yablo's Game**, in this case, is that E cannot come up with *any acceptable revision sequence* for any sentence in **Yablo's Game** *at any point*, as the above reasoning shows. Condition (II) serves the only purpose to exclude "lucky" assignments of semantic values – e.g. the scenario where E guesses that P is producing a Yablo sequence (this is sheer luck, as E does not have any indication in this sense) and, adopting for example the evaluation given by Kripke least fixed point plus C-gaps, she assigns value $1/2$ to every sentence P utters. Any such scenario is not relevant in understanding how semantic values are assigned or should be assigned.

Sentence Y seems to provide the piece of information recalled above, that is present in Yablo's semi-formal version of the paradox and in the formalized versions too, enabling *E* to determine that *P* is producing a Yablo sequence and to win the game. But this is not so: in fact, one could run the above proof of the impossibility to win **Yablo's Game** even if *P* utters *Sentence Y*.[33] In order for *E*'s situation to be different, she would need to know that *Sentence Y holds*, i.e. that *P* will act according to what *Sentence Y* says. But this information cannot come from within a game where one player must evaluate the other's sentences! Any information that could put *E* in a situation to win **Yablo's Game** has to come from outside of the game, i.e., leaving fiction aside, outside of the speakers interpreting and evaluating their language.[34] The relations holding between the sentences in **Yablo's Game** can only be identified if one takes into account information that *cannot come* from within the game.

Moreover, even if one might argue (with Priest) that every sentence in **Yablo's Game** "says the same thing", there is no structural feature of those sentences that determines the nature of the entire list. The sentences of **Yablo's Game** are perfectly innocuous in themselves, as the following game shows.

Trivially Simple Game.

> "Call this very sentence "*Sentence k*": all the sentences I will utter after *Sentence k* are not true".
>
> "0=0". [End of Game]

E can win the **Trivially Simple Game**. So, there is nothing special in the sentences of **Yablo's Game** – in this respect, they are similar to contingent liars.

[33] Only item 2 would require a small modification.

[34] One might object that at least one speaker can determine whether what *P* says is true, false, or something else, namely *P*. This is not directly relevant to the problems of evaluating sentences, which is embodied by *E*'s role in the game; however, we can imagine a scenario where *P* follows some random process in order to decide which sentences to utter and where, accidentally, she produces larger and larger fragments of a Yablo list.

Cases such as **Yablo's Game** are completely different from cases that require to assign a semantic value to infinitely many sentences but where such assignment is possible *at some point*, for an idealized agent. Contrast **Yablo's Game** with, e.g., the following natural language version of McGee's sentence.

(MG) It is not the case that the sentence resulting from prefixing the sentence called "(MG)" with any number of instances of the truth predicate is true.

An idealized agent, after considering all the infinitely many components of (MG), can assign a value to (MG). The same idealized agent, however, is never in a position to evaluate safely a sentence of **Yablo's Game**.

The relevance of **Yablo's Game** to understand and model simple languages should not be underestimated. This version of Yablo's paradox does not require infinitely many sentences (unlike Yablo's original sequence), nor infinitely long sentences (unlike [Cook 2006]'s variant): **Yablo's Game** only requires that the sentences in it are produced so long as there are people that understand English.[35] So, despite the obvious artificiality of **Yablo's Game**, its upshot is clear: some simple sentences (possibly formulated in a language featuring only mathematical, syntactic, and semantic vocabulary) *should not* be assigned any semantic value at all, and not because of some unknown mathematical or scientific facts.[36]

Yablo's Game is just a paradigmatic case, but a similar situation is given whenever one confronts a set of sentences that do not offer a starting point to work one's

[35]This could be accomplished having always new people acting as Evaluators (as the older ones quit) and replacing P with some computer program that, unbeknownst to all, produces a larger and larger fragment of a Yablo list, ensuring that the program (or its replacements) keeps doing its job at least until there are people understanding English.

[36]This conclusion does not exclude the thesis that sentences that should not be assigned a value, as those of **Yablo's Game**, *have* a semantic value nonetheless. This issue seems hard to settle but is completely independent from the assignability of semantic values, so we can set it aside without consequences for the present work. The fact that the sentences in **Yablo's Game** *admit a consistent assignment* of semantic values, as mentioned in footnote 32, does not show that they *have* a semantic value. Personally, I believe that an account of semantic values that holds that the sentences in **Yablo's Game** have a semantic value is very unlikely to find a convincing justification, but the semantics I propose is officially agnostic on these matters.

way towards an assignment of values. Such cases show that the standard compositional approaches to evaluations[37] should be refined to take into account whether the sentences to be evaluated, their components, and their structures admit a starting point at which semantic values can be safely assigned. This, in turn, vindicates (IAC) as a satisfactory compositional account of the ideal semantic process that leads to evaluate the sentences of the simple languages targeted in this work.

(IAC) is embodied by the canonical evaluation, that gives no value to Yablo-like sentences. The canonical evaluation treats Yablo sentences from a purely semantic point of view (i.e. without taking into account any other sort of information such as their fixed-points properties). From this point of view, every Yablo sentence generates always different sentences as components, and the inspection of the latter never provides us a starting point to assign them a semantic value. So, for example, when one confronts the k-th Yablo sentence, she confronts a sentence of the form $\forall n > k \neg Tt$, no more and no less. To find out something about this sentence *semantically* and *compositionally*, we should consider its components, if it has any. This is the case, and we should now consider the instances of the universal quantifier. Doing so, more sentences are revealed, and so on *ad infinitum*. At *no point* in this process we can identify the structure common to all the sentences ahead of us, looking only at the semantics of each Yablo sentence. **Yablo's Game** converts this situation in a plausible, natural language case that makes it clear why we should accept (IAC) and the formal semantics that embody it, such as the canonical evaluation. (IAC) and cases such as **Yablo's Game** entail that we should not assign a semantic value to some sentences: this is a substantial conclusion in line with the main tasks set forth for this work, namely interpreting at best some simple languages including mathematical and semantic vocabulary, and treating satisfactorily the paradoxes that affect them.

[37] Namely, those views according to which a semantic value is assigned to a sentence based on the assignments of values to their components.

6.4.3 Lawlessness

If one accepts a compositional view such as (IAC), and the consequent treatment of Yablo-like cases, she should also accept that some instances of every classical law are not assigned any semantic value, and in particular that they are not assigned the/a designated value. *Prima facie*, such "lawlessness" seems an undesirable feature of a semantic theory: now I will try to argue for the contrary.

Let me start by noticing that lawlessness is not a strange position in the context of theories of naïve truth. As there are many different theories of naïve truth validating different (and mutually incompatible) sets of laws, it is unclear that we should attribute a privileged status to any such set, without independent and conclusive reasons to do so. Lawlessness appears a neutral ground in this respect.[38]

In addition, as mentioned, lawlessness flows quite naturally from a compositional view of semantic value attributions: if φ receives no value at all and we accept a compositional picture of semantic value assignments, it is natural that $\varphi \vee \neg\varphi$, $\varphi \rightarrow \varphi$, and similar laws involving φ should receive no value as well. Now, since I have argued that we should not assign semantic values to some sentences, we should refrain also from assigning semantic values to instances of classical laws that feature such sentences. For example, **Yablo's Game** would not be any less problematic if P uttered, amongst the other sentences, a sentence as the following: "if *Sentence m* is true then *Sentence m* is true".

Finally, some independently motivated views can be used to support lawlessness. For example, in the ground-theoretic literature we find the view according to which "any necessary truth [is] ultimately to be grounded in contingent truths".[39] This view is often cashed out in terms of classical laws, that are seen as paradigmatic cases of necessary truth. Ground-theoretic debates are not directly relevant

[38] This aspect was discussed, in conjunction with Field's theory, in Subsection 3.5.4. See also [Halbach 2011], p. 245 and following and p. 292 and following.

[39] This idea is mentioned in [Fine 2012], p. 49.

to my concerns, but they show that there are well-developed views on the status of logical laws that reduce their semantic value to worldly matters.[40] Such views may be used just as well for the more modest purpose of reducing the (idealized) *assignment* of semantic values to worldly matters: cases such as **Yablo's Game** can be used to justify the choice of not assigning semantic values in some circumstances and, together with a compositional view such as (IAC), to justify lawlessness. Adopting this view, if we take the above example where P utters also "if *Sentence m* is true then *Sentence m* is true" at some point in **Yablo's Game**, we should conclude that the semantic value assigned to this statement is grounded in the semantic value assigned to *Sentence m* (via naïveté). As we should not assign any semantic value to the latter sentence, we should not assign any semantic value to "if *Sentence m* is true then *Sentence m* is true" either.

6.5 The numerical values

The canonical evaluation gives a simple way to understand numerical values: when φ receives value k in V, the number k corresponds to the *only* way to interpret φ numerically, given our construal of connectives, quantifiers, and T.

An advantage of numerical values is their simplicity. In Field's theory, e.g., the semantic values themselves are complex objects, quite difficult to make sense of.[41]

A common objection to many-valued semantics employing many numerical values is that some value assignments seem absurd or arbitrary.[42] For example, doesn't it seem absurd or arbitrary to give ψ semantic value $7/2^{57885161}-1$?[43] Given the enormity of the denominator, it is difficult to have an intuitive idea of the size of the fraction. Moreover, since the value in question is so small, what difference

[40] Whether such matters are facts, states of affairs, or else is of no importance here.
[41] See Definition 3.3.5 and the subsequent discussion.
[42] See for example [Williamson 1994], Chapter 4.
[43] While I am writing, $2^{57885161-1}$ is the largest known prime number.

would it make if we assigned value $^{6}/_{2^{57885161}-1}$ to ψ? I do not know whether some $\mathscr{L}_T^{\rightarrow}$-sentence is assigned value $^{7}/_{2^{57885161}-1}$ by the canonical evaluation but, even if it is, there is no absurdity or arbitrariness in such assignment. In fact, there is no arbitrariness or absurdity in *any* numerical assignment: a numerical value is always the *only* way to construe numerically a sentence in V, respecting the chosen meaning of connectives, quantifiers, and T. So, any absurdity or arbitrariness one finds in the numerical results of the canonical evaluation should be attributed to the interpretation one gives to connectives, quantifiers, and T. However, the chosen algebraic interpretations of connectives, quantifiers, and T are quite common and do not seem absurd or arbitrary at all, and numerical evaluations are just natural ways to implement algebraic interpretations.

These observations, however, may be considered an unsatisfactory justification for the choice of using more than 2 values, since they do not answer a *vexata quaestio* in the area of many-valued logics: *what do the extra semantic values mean*? While there is some consensus that 1 and 0 coincide with or approximate the classical semantic values, it is not so clear how to understand the intermediate values. $^{1}/_{2}$ is an exception, to some extent, in 3-valued logics, because it can be construed as embodying an alternative to classical semantic values. But what is the intuitive understanding of semantic value $^{7}/_{31}$?

I do not think, however, that authors adopting multiple numerical values should feel compelled to address this issue. Many theorists are mainly interested in sentences having the designated value(s), and certainly not because they are interested in those values, but only because when a sentence has the/a designated value, one is presumably entitled to accept and assert it. If one adopts the canonical evaluation and wishes to remain agnostic as to real "significance" or "nature" of the numerical values, she can easily do that. Although I argued that the fact that a sentence is assigned a value in Ł or no value at all by the canonical evaluation does reveal important features of such sentence, it seems a virtue of this model that it does not force nor suggest a specific reading for the numerical values, leaving one free to choose any view she likes. Finally, note that this does not belittle the importance of the extra numerical values, nor their features (such as the cardinality of V, or the fact that V is a totally ordered set). After all, it is thanks to the presence and the role of intermediate values interpreted in a set such as V that one can justify several philosophically non-trivial claims based on 1-valued sentences, such as "the liar is equivalent to its negation and to Curry's sentence", or "some sentences are less determinate than others by arithmetical facts and by the meaning of connectives, quantifiers, and truth", and many more.

6.6 Summary of the chapter

In this chapter, I argued for the canonical evaluation, as providing a satisfactory semantics for $\mathscr{L}_T^{\rightarrow}$ and an acceptable treatment of truth-theoretic paradoxes.

After a quick introduction (Section 6.1), I showed how the canonical evaluation addresses FIELD'S PROGRAM in Section 6.2. Section 6.3 explored the use of Łukasiewicz equations as semantic values in treating paradoxes, analyzing equation systems with no solutions (Subsection 6.3.1) and equation systems with multiple solutions (Subsection 6.3.2). Section 6.4 addressed some issues related to Yablo-like paradoxes, namely: circularity and self-reference (Subsection 6.4.1), a compositional view on the assignment of semantic values (Subsection 6.4.2), and some justification for the fact that no classically valid logical law is validated by the canonical evaluation (Subsection 6.4.3). Finally, in Section 6.5 I discussed the status of numerical semantic values in the canonical evaluation.

Bibliography

Aczel, Peter (1977) “An Introduction to Inductive Definitions”, in J. Barwise (ed.), *Handbook of Mathematical Logic*, North-Holland, Amsterdam: 739-782.

Aczel, Peter and Solomon Feferman (1980), “Consistency of the unrestricted abstraction principle using an intensional equivalence operator”, in J.P. Seldin and J.R. Hindley (eds.), *To H. B. Curry: Essays on combinatory logic, lambda calculus and formalism*, Academic Press, New York: 67-98.

Bacon, Andrew (2013a), “A New Conditional for Naive Truth Theory”, *Notre Dame Journal of Formal Logic*, 54, 1: 87-104.

Bacon, Andrew (2013b), “Non-Classical Metatheory for Non-Classical Logics”, *Journal of Philosophical Logic*, 42, 2: 335-355.

Barwise, Jon and John Etchemendy (1987), *The Liar. An Essay on Truth and Circularity*, Oxford University Press, Oxford.

Beall, JC (2001), “Is Yablo’s Paradox Non-Circular?”, *Analysis*, 61: 176-87.

Beall, JC (ed.) (2003), *Liars and Heaps. New Essays on Paradox*, Oxford University Press, Oxford.

Beall, JC and Brad Armour-Garb (eds.) (2004), *Deflationism and Paradox*, Oxford University Press, Oxford.

Beall, JC (ed.) (2007), *Revenge of the Liar. New Essays on the Paradox*, Oxford University Press, Oxford.

Beall, JC (2009), *Spandrels of Truth*, Oxford University Press, Oxford.

Belnap, Nuel D. Jr. (1977), "A useful four-valued logic", in J. M. Dunn and G. Epstein (eds.), *Modern Uses of Multiple-Valued Logic*, Reidel, Dordrecht: 8-37.

Belnap, Nuel and Anil Gupta (1993), *The Revision Theory of Truth*, MIT Press, Cambridge (MA).

Blamey, Stephen (2002), "Partial Logic", in [Gabbay and Guenthner 2002], volume V: 261-353.

Bondy, Adrian and U.S.R. Murty (2008), *Graph Theory*, Springer, London.

Brady, Ross (1971), "The Consistency of the Axioms of Abstraction and Extensionality in Three-Valued Logic", *Notre Dame Journal of Formal Logic*, 12, 4: 447-453.

Bueno, Otávio and Mark Colyvan (2003), "Paradox Without Satisfaction", *Analysis*, 63: 152-56.

Burgess, John P. (1986), "The Truth is Never Simple", *Journal of Symbolic Logic*, 51, 3: 663-681.

Burgess, John P. (2014), "Friedman and the Axiomatization of Kripke's Theory of Truth", in N. Tennant (ed.), *Foundational Adventures. Essays in Honour of Harvey Friedman*, College Publications.

Cain, James and Zlatan Damnjanovic (1991), "On the Weak Kleene Scheme in Kripke's Theory of Truth", *Journal of Symbolic Logic*, 56, 4: 1452-1468.

Cantini, Andrea (1990), “A Theory of Formal Truth Arithmetically Equivalent to ID_1”, *Journal of Symbolic Logic*, 55, 1: 244-259.

Cantini, Andrea (1996), *Logical Frameworks for Truth and Abstraction. An Axiomatic Study*, Elsevier, Amsterdam.

Cantini, Andrea (2002), “Osservazioni su Autoriferimento e Verità”, *Annali del Dipartimento di Filosofia dell’Università degli Studi di Firenze*, 8: 53-76.

Cantini, Andrea (2009), “Paradoxes, Self-Reference and Truth in the Twentieth Century”, in D. Gabbay and J. Woods (eds.), *Handbook of the History of Logic*, Elsevier, Amsterdam, volume V: 875-1013.

Cook, Roy T. (2004), “Pattern of Paradox”, *Journal of Symbolic Logic*, 69, 3: 767-774.

Cook, Roy T. (2006), “There Are Non-circular Paradoxes (But Yablo’s Isn’t One of Them!)”, *The Monist*, 89, 1: 118-149.

Cook, Roy T. (2014), *The Yablo Paradox: An Essay on Circularity*, Oxford University Press, Oxford.

Davis, Lawrence H. (1979), “An Alternate Formulation of Kripke’s Theory of Truth”, *Journal of Philosophical Logic*, 8, 1: 289-296.

Diestel, Reinhardt (2010), *Graph Theory*, fourth edition, Springer, Berlin.

Feferman, Solomon (1984), “Toward Useful Type-Free Theories. I”, *Journal of Symbolic Logic*, 49, 1: 75-111.

Feferman, Solomon (1991), “Reflecting on Incompleteness”, *Journal of Symbolic Logic*, 56, 1: 1-49.

Field, Hartry (1994a), “Deflationist View of Meaning and Content”, *Mind*, 103, 411: 249-285.

Field, Hartry (1994b), “Disquotational Truth and Factually Defective Discourse”, *The Philosophical Review*, 103, 3: 405-452.

Field, Hartry (1999), “Deflating the Conservativeness Argument”, *Journal of Philosophy*, 96, 10: 541-544.

Field, Hartry (2002), “Saving the Truth Schema from Paradox”, *Journal of Philosophical Logic*, 31, 1: 1-27.

Field, Hartry (2003a), “A Revenge-Immune Solution to the Semantic Paradoxes”, *Journal of Philosophical Logic*, 32, 2: 139-177.

Field, Hartry (2003b), “The Semantic Paradoxes and the Paradoxes of Vagueness”, in [Beall 2003]: 262-311.

Field, Hartry (2004a), “The Consistency of the Naïve Theory of Properties”, *The Philosophical Quarterly*, 54, 214: 78-104.

Field, Hartry (2004b), “Is the Liar Sentence Both True and False?”, in [Beall and Armour-Garb 2004]: 23-40.

Field, Hartry (2004c), “Variations on a Theme by Yablo”, in [Beall and Armour-Garb 2004]: 53-73.

Field, Hartry (2007), “Solving the Paradoxes, Escaping Revenge”, in [Beall 2007]: 78-144.

Field, Hartry (2008), *Saving truth from paradox*, Oxford University Press, Oxford.

Field, Hartry (2011), “Comments on Martin’s and Welch’s Comments”, *Review of Symbolic Logic*, 4, 3: 360-366.

Field, Hartry (2014), “Naive Truth And Restricted Quantification: Saving Truth A Whole Lot Better”, *Review of Symbolic Logic*, 7: 147-91.

Field, Hartry, Harvey Lederman and Tore Fjetland Øgaard (forthcoming), “Prospects for a Naïve Theory of Classes”, to appear in the *Notre Dame Journal of Formal Logic*.

Fine, Kit (2012), “Guide to Ground”, in F. Correia and B. Schnieder (eds.), *Metaphysical Grounding*, Cambridge University Press, Cambridge: 37-80.

Fischer, Martin, Volker Halbach, Jönne Speck, Johannes Stern (2015), “Axiomatizing Semantic Theories of Truth?”, *Review of Symbolic Logic*, 8: 257-278.

Fitting, Melvin (1986), “Notes on the Mathematical Aspects of Kripke’s Theory of Truth”, *Notre Dame Journal of Formal Logic*, 27, 1: 75-88.

Gaifman, Haim (1988), “Operational Pointer Semantics: Solution to Self-referential Puzzles I" in M. Vardi (ed.) *Theoretical Aspects of Reasoning about Knowledge* Morgan Kauffman, Los Angeles: 43-59.

Gaifman, Haim (1992), “Pointers to Truth”, *Journal of Philosophy*, 89, 5: 223-261.

Gaifman, Haim (2000), “Pointers to Propositions”, in A. Chapuis and A. Gupta (eds.), *Circularity, Definition and Truth*, Indian Council of Philosophical Research, New Delhi.

Gabbay, Dov M. and Franz Guenthner (eds.) (2002), *Handbook of Philosophical Logic*, second edition, Kluwer, Dordrecht.

Gilmore, Paul C. (1967), “The Consistency of a Positive Set Theory”, IBM Research Report RC-1754, January 24.

Gilmore, Paul C. (1974), "The Consistency of Partial Set Theory without Extensionality", in T. Jech (ed.) *Axiomatic set theory*, American Mathematical Society, Providence: 147-153.

Gottwald, Siegfried (2001), *A Treatise on Many-Valued Logics*, Research Studies Press, Baldock.

Hájek, Petr and Pavel Pudlák (1998), *Metamathematics of First-Order Arithmetic*, Springer, Berlin.

Hájek, Petr, Jeff Paris, and John Shepherdson (2000), "The Liar Paradox and Fuzzy Logic", *Journal of Symbolic Logic*, 65, 1: 339-346.

Halbach, Volker (1996), "Tarski-hierarchies", *Erkenntnis*, 43: 339-367.

Halbach, Volker (1999), "Disquotationalism and Infinite Conjunctions", *Mind*, 108, 429: 1-22.

Halbach, Volker (2001), "Disquotational Truth and Analyticity", *Journal of Symbolic Logic*, 66, 4: 1959-1973.

Halbach, Volker (2001b), "How innocent is deflationism?", *Synthese*, 26: 167-194.

Halbach, Volker (2002), "Modalized Disquotationalism", in V. Halbach and L. Horsten (eds.), *Principles of Truth*, Hänsel-Hohenhaus, Frankfurt: 75-101.

Halbach, Volker, Hannes Leitgeb, and Philip Welch (2005), "Possible Worlds Semantics for Predicates", in R. Kahle (ed.) *Intensionality*, A K Peters, Wellesley: 20-41.

Halbach, Volker and Leon Horsten (2006a), "Axiomatizing Kripke's Theory of Truth", *Journal of Symbolic Logic*, 71, 2: 677-712.

Halbach, Volker (2006b), “How not to state the T-sentences”, *Analysis*, 66: 276-280. Correction of printing error in vol. 67: 268.

Halbach, Volker and Philip Welch (2009), “Necessities and Necessary Truths: A Prolegomenon to the Metaphysics of Modality”, *Mind* 118: 71-100.

Halbach, Volker (2011), *Axiomatic Theories of Truth*, Cambridge University Press, Cambridge.

Halbach, Volker and Albert Visser (2014a), “Self-Reference in Arithmetic I”, *Review of Symbolic Logic*, 7, 4: 671-691.

Halbach, Volker and Albert Visser (2014b), “Self-Reference in Arithmetic II”, *Review of Symbolic Logic*, 7, 4: 692 - 712.

Hänle, Reinhardt (2002), “Advanced Many-Valued Logic”, in [Gabbay and Guenthner 2002], volume II: 297-395.

Hansen, Casper Storm (2015), “Supervaluation on Trees for Kripke’s Theory of Truth”, *Review of Symbolic Logic*, 8: 46-74.

Hazen, Allen (1981) “Davis’s formulation of Kripke’s theory of truth: A correction”, *Journal of Philosophical Logic*, 10, 3: 309-311.

Hodges, Wilfrid (1993), *Model Theory*, Cambridge University Press, Cambridge.

Horsten, Leon (1998), “A Kripkean Approach to Unknowability and Truth”, *Notre Dame Journal of Formal Logic*, 39, 3: 389-405.

Horsten, Leon (2003), “The Logic of Intensional Predicates”, in B. Löwe, W. Malzkorn, and T. Räsch (eds.) *Foundations of the Formal Sciences II*, Kluwer Academic, Dordrecht: 89-111.

Horsten, Leon (2009), “Levity”, *Mind*, 118: 555-581.

Horsten, Leon (2011), *The Tarskian Turn*, MIT Press.

Horwich, Paul (1998), *Truth*, second edition, Basil Blackwell, Oxford.

Jech, Thomas (2002), *Set Theory*, third millennium edition, Springer.

Kaye, Richard (1991), *Models of Peano Arithmetic*, Clarendon Press, Oxford.

Ketland, Jeffrey (1999), “Deflationism and Tarski’s Paradise” *Mind*, 108, 429: 165-72.

Ketland, Jeffrey (2004), “Bueno and Colyvan on Yablo’s Paradox” *Analysis*, 64: 165-72.

Kleene, Stephen Cole (1938), “On Notation for Ordinal Numbers”, *Journal of Symbolic Logic*, 3, 4: 150-155.

Kleene, Stephen Cole (1952), *Introduction to Metamathematics*, Van Nostrand, New York.

Kremer, Michael (1988), “Kripke and the Logic of Truth”, *Journal of Philosophical Logic*, 17, 3: 327-332.

Kripke, Saul (1975), “Outline of a theory of truth”, *Journal of Philosophy*, 72, 19: 690-716.

Leitgeb, Hannes (2002), “What is a Self-referential Sentence? Critical Remarks on the Alleged (non-)Circularity of Yablo’s Paradox”, *Logique et Analyse* 177-178: 3-14.

Leitgeb, Hannes (2005), “What Truth Depends On”, *Journal of Philosophical Logic*, 34: 155-192.

Leitgeb, Hannes (2006), “Towards a Logic of Type-Free Modality and Truth” in C. Dimitracopoulos, L. Newelski, D. Normann, J. R. Steel (eds.) *Logic Colloquium 2005*, Cambridge University Press, Cambridge.

Leitgeb, Hannes (2007), “On the Metatheory of Field’s “Solving the Paradoxes, Escaping Revenge”, in [Beall 2007]: 159-183.

Malinowski, Grzegorz (1993), *Many-Valued Logics*, Clarendon Press, Oxford.

Martin, Robert L. and Peter Woodruff (1975), “On Representing ‘True-in-*L*’ in *L*”, *Philosophia*, 5: 213-217.

Martin, Donald A. (2011), “Field’s *Saving Truth from Paradox*: Some Things It Doesn’t Do”, *Review of Symbolic Logic*, 4, 3: 339-347.

Maudlin, Tim (2004), *Truth and Paradox*, Oxford University Press, New York.

McGee, Vann (1985), “How Truthlike can a Predicate be? A Negative Result”, *Journal of Philosophical Logic*, 14, 4: 399-410.

McGee, Vann (1989), “Applying Kripke’s Theory of Truth”, *Journal of Philosophy*, 86, 10: 530-539.

McGee, Vann (1991) *Truth, Vagueness, and Paradox: An Essay on the Logic of Truth*, Hackett, Indianapolis.

McGee, Vann (1992), “Maximal Consistent Sets of Instances of Tarski’s Schema (T)”, *Journal of Philosophical Logic*, 21, 3: 235-241.

Meadows, Toby (2013), “Truth, Dependence and Supervaluation: Living with the Ghost”, *Journal of Philosophical Logic*, 42: 221-240.

Minari, Pierluigi (1987), Unpublished notes, Florence.

Moschovakis, Yiannis N. (1974), *Elementary Induction on Abstract Structures*, North-Holland, republication by Dover Books, New York (2008).

Pailos, Federico and Lucas Rosenblatt (2015), "Non-Deterministic Conditionals and Transparent Truth", *Studia Logica*, 103, 3: 579-598.

Priest, Graham (1979), "Logic of Paradox", *Journal of Philosophical Logic*, 8: 219-241.

Priest, Graham (1990), "Boolean negation and all that", *Journal of Philosophical Logic* 19: 201-15.

Priest, Graham (1997), "Yablo's Paradox", *Analysis* 57: 236-42.

Priest, Graham (2006), *In Contradiction. A Study of the Transconsistent*, second edition, Oxford University Press, Oxford.

Pohlers, Wolfram (2009), *Proof Theory. The First Step into Impredicativity*, Springer, Berlin.

Rabern, Landon, Brian Rabern, and Matthew Macauley (2013), "Dangerous reference graphs and semantic paradoxes", *Journal of Philosophical Logic*, 42, 5: 727-765.

Rasiowa, Helena (1974), *An Algebraic Approach to Non-Classical Logics*, North Holland, Amsterdam.

Restall, Greg (1992), "Arithmetic and Truth in Łukasiewicz's Infinitely Valued Logic", *Logique Et Analyse*, 139 (140): 303-312.

Rogers, Hartley Jr. (1987), *Theory of Recursive Functions and Effective Computability*, second edition, MIT Press, Cambridge (MA).

Scharp, Kevin (2007), "Alethic Vengeance", in [Beall 2007]: 272-319.

Schlenker, Philippe (2007), “The Elimination of Self-Reference: Generalized Yablo-Series and the Theory of Truth”, *Journal of Philosophical Logic*, 36, 3: 251-307.

Shapiro, Stewart (1998), “Proof and Truth: Through Thick and Thin”, *Journal of Philosophy*, 95, 10: 493-521.

Sheard, Michael (1994), “A guide to truth predicates in the modern era”, *Journal of Symbolic Logic*, 59, 3: 1032-1054.

Smiley, Timothy (1962), “The Independence of Connectives”, *Journal of Symbolic Logic*, 27, 4: 426-436.

Soames, Scott (1999), *Understanding Truth*, Oxford University Press, Oxford.

Sorensen, Roy A. (1998), “Yablo’s Paradox and Kindred Infinite Liars”, *Mind*, 107, 425: 137-155.

Szabó, Zoltán Gendler (2000), *Problems of Compositionality*, Garland, New York.

Tarski, Alfred (1935), “The Concept of Truth in Formalized Languages”, English translation in Alfred Tarski (1983), *Logic, Semantics, Metamathematics*, second edition, edited by J. Corcoran, Hackett, Indianapolis.

Urquhart, Alasdair (2002), “Basic Many-Valued Logic”, in [Gabbay and Guenthner 2002], volume II: 249-295.

van Dalen, Dirk (2012), *Logic and Structure*, fifth edition, Springer, London.

Visser, Albert (2002), “Semantics and the Liar Paradox”, in [Gabbay and Guenthner 2002], volume XI: 149-240.

Welch, Philip (2009), “Ultimate truth *vis à vis* stable truth”, *Review of Symbolic Logic*, 1, 1: 126-142.

Welch, Philip (2011), “Truth, Logical Validity and Determinateness: A Commentary on Field’s *Saving Truth from Paradox*”, *Review of Symbolic Logic* 4, 3: 248-259.

Welch, Philip (2014), “Some Observations on Truth Hierarchies”, *Review of Symbolic Logic*, 7, 1: 1-30.

Whittle, Bruno (2015), “Truth, Hierarchy and Incoherence”, to appear in B. Armour-Garb (ed.), *The Relevance of the Liar* (forthcoming).

Williamson, Timothy (1994), *Vagueness*, Routledge, London.

Yablo, Stephen (1982), “Grounding, Dependence, and Paradox” *Journal of Philosophical Logic*, 11: 117-137.

Yablo, Stephen (1985), “Truth and Reflection”, *Journal of Philosophical Logic*, 14: 279-349.

Yablo, Stephen (1993), “Paradox Without Self-Reference”, *Analysis*, 53: 251-52.

Yablo, Stephen (2003), “New Grounds for Naive Truth Theory”, in [Beall 2003]: 312-330.

Yablo, Stephen (2006), “Circularity and Paradox”, in T. Bolander, V. Hendricks, S. Pedersen (eds.), *Self-Reference*, CSLI Publications, Stanford: 139-157.

Yi, Byeong-uk (1999), “Descending Chains and the Contextualist Approach to Paradoxes”, *Notre Dame Journal of Formal Logic*, 40, 4: 554-567.

Zardini, Elia (2011), “Truth without Contra(di)ction”, *Review of Symbolic Logic*, 4: 498-535.

Zardini, Elia (2013), “Naive *Modus Ponens*”, *Journal of Philosophical Logic*, 42: 575-593.