Truth in *FDE*

Ane Maria G. Døhl

Thesis presented for the degree of

MASTER OF PHILOSOPHY

Supervised by Professor Øystein Linnebo

Department of Philosophy, Classics, History of Art and Ideas

UNIVERSITY OF OSLO

December 2022

Acknowledgments

First, I want to thank my supervisor Øystein Linnebo for his guidance, encouragement and for always explaining complex issues in a simple manner.

Second, I would like to thank the Centre for Philosophy and the Sciences for their generous support of my thesis. In particular, I want to thank Gry Oftedal for inviting me to present my thesis at one of the CPS lunch meetings. I would also like to thank the organisers of the Summer School of the Munich Centre for Mathematical Philosophy Summer school for letting me present this thesis. I am very grateful for all the participants for their valuable feedback.

Third, I want to thank those who have commented on the thesis along the way, in particular Zihao Guo and Alejandro Solares-Rojas.

I also owe a debt of gratitude to my fellow students for their support during these two years. Doing a degree during a pandemic has been a particular experience, and I want to extend my gratitude especially to the Tuesday coffee group: Andrea Foss, Isabella Bartoli, Christer Hølestøl and Amalie Torkelsen Friis.

Finally, I would like to thank my family who have helped and encouraged me during these two years. In particular, I want to express my gratitude and love to my parents Hannah Gerdes and Jon Døhl.

Abstract

This thesis explores the concept of truth in the logical system First-Degree Entailment FDE, attempting to balance the ordinary conception of truth on the one hand, with the need to provide a solution to the liar paradox on the other. Starting from Tarski's tripartite analysis of the liar paradox as being caused by self-reference, the T-schema and classical logic, this thesis argues that the two first elements should be kept. Instead, classical logic is abandoned and the thesis explores non-classical solutions. It presents paracomplete and paraconsistent solutions to the liar paradox, before defending a unifying framework, namely the logic FDE. Finally, it addresses the issue of the revenge paradox and sketches out two possible responses.

Contents

Preface

I began working on this project two years ago, in the middle of a pandemic, amidst worrying disinformation campaigns and ludicrous conspiracy theories. My main interest within philosophy has always been logic, a field where the concept of truth is central. There seemed to be a chasm, however, between the various truth theories in academia and the difficulties surrounding truth in society. What good is a theory, no matter how elegant, if it presents a truth predicate which I cannot recognise as truth? The starting point of this project was therefore a small attempt to bridge the gap between logical theories and a more intuitive conception of truth. The thesis has since evolved, but maintaining a balance between intuition and ordinary speech on the one hand, and technical concerns on the other hand, has remained a guiding light.

My other starting point was the liar paradox and its Tarskian tripartite diagnosis. According to Tarski, the liar paradox is caused by self-reference, the T-schema and classical logic. I argue that for truth to be recognisable as our truth, we need to keep self-reference and the T-schema. This allowed me to explore truth in non-classical logic, and in particular the logic of FDE.

Chapter 1 is an extended introduction, where I set up the liar paradox and Tarski's analysis in more detail. In the last section I say a little more about the general methodology of the thesis and the concern of finding an equilibrium between what we would like a truth theory to be like, and how the concept of truth is in fact understood. In the next two chapters I defend keeping self-reference (chapter 2) and the T-schema (chapter 3).

The exploration of truth in non-classical logic starts from chapter 4, in which I present the paracomplete and paraconsistent solutions to the liar paradox. In chapter 5 I compare the two solutions, and argue that instead of arguing for one over the other, the logic FDE provides a unifying framework. This logic is presented in chapter 6, where I argue for a specific implication connective in order to formulate the T-schema in FDE. In the last chapter, I set up a revenge paradox for truth in FDE and sketch out a few solutions.

Chapter 1 – Introduction

Section 1 – The liar paradox

Aase: Hvad? Har du løjet nu igjen?
Peer Gynt: Ja, denne Gang.
Peer Gynt, I,i[1]

"I lie", says Peer Gynt, after having told a story about his fight with the blacksmith Aslak. But can we believe him? Is he telling the truth about having just lied? Or, inveterate liar that he is, is he again lying, and this time lying about lying? If Peer is telling the truth, then it is true that he lies, in which case "I lie" is a lie and Peer is not telling the truth. But if Peer Gynt is lying, then "I lie" is itself a lie and Peer is in fact telling the truth.

What is the truth-value of Peer Gynt's statement? If it is true then it is false, and if it is false then it is a lie. This is the famous liar paradox, which has been a much-discussed topic in the West since at least the Ancient Greeks.[2]

The easiest way to state the paradox is as follows:

$$\lambda \coloneqq \lambda \ is \ false$$

The sentence λ says of itself that it is false. If λ is true, then 'λ is false' is true, that is, λ is false. But if λ is false, then 'λ is false' is false, that is, λ is true.

The liar paradox can take many shapes, for example with the help of loops and cycles.

Peer Gynt: "Pinocchio is lying"

Pinocchio: "Peer Gynt is telling the truth"

If Peer tells the truth, then Pinocchio lies about Peer telling the truth, which means that Peer lies. But if Peer lies, Pinocchio is telling the truth and Peer is not lying.

An infinity of more complex variations can be imagined, but the core issue of the liar paradox remains. This paradox has been extremely resilient and near-impossible to solve. Any proposed

[1] ÅSE: What? You're lying now again?
PEER: Yes, just this once.
Ibsen, *Peer Gynt*.

[2] Beall, Jc, Michael Glanzberg, and David Ripley, "Liar Paradox".

solution has yet failed to reach a broad consensus. Even worse is that the liar paradox sometimes appears in different forms to cause further mischief. A modified version of the liar sentence, 'this sentence is not provable', is at the core of Gödel's first incompleteness theorem, which states that any formal system strong enough to carry out elementary arithmetic cannot be both complete and consistent. Either there will be true statements that cannot be proved, or the system will prove false statements.

Paradoxes are often considered guides to philosophical inquiry, and have been used to question many ordinary concepts such as time, identity, truth or knowledge. The liar paradox has thus sparked a lot of discussion about the nature of truth, and has been seen as revealing a flaw within the concept of truth itself. It is therefore a natural starting point when examining the nature of truth.

Section 2 – The recipe for the liar paradox: Tarski's analysis

Alfred Tarski used the liar paradox as a starting point for his inquiry into the truth predicate. In the first section of "The concept of truth in formalized languages"[3] he analyses the liar paradox as being caused by three features: the language must be capable of expressing self-reference, the truth predicate must be based on the T-schema and the language must be able to formulate the liar sentence. A fourth implicit feature is classical logic. Any language with sentences that can express their own truth or falsity, such as natural language, cannot have a concept of truth based on the T-schema that is consistent in classical logic. Tarski then sets out his own truth predicate for formal languages, designed specifically so as to avoid the liar paradox.

Where Tarski decides to keep the T-schema and classical logic and give up natural language, this thesis aims at exploring what happens if one keeps the T-schema and natural language, in particular its ability to have self-referential sentences expressing their own truth-value, but gives up classical logic instead.

This section will give an overview of Tarski's analysis before setting out the strategy of this thesis in more detail.

[3] Tarski, "The concept of truth in formalized languages".

2.1 Tarski's analysis

Tarski himself does not explicitly list self-reference, the T-schema and classical logic as the three elements for the liar paradox. Instead, he lists self-reference, the language's ability to state the liar paradox and the T-schema. Classical logic is taken as an implicit self-evident premise.

2.1.1 First attempt to define truth: the semantic definition

Tarski first attempts to define truth by defining what a true sentence is. His goal is to provide a precise articulation of the relationship between what a true sentence says and the "state of affairs" in the world, but using only clear and precisely defined terms.

> (1) a true sentence is one which says that the state of affairs is so and so, and the state of affairs indeed is so and so.[4]

This definition is too vague and imprecise, so Tarski attempts to refine it by giving the following formula, in which p can be replaced by any sentence and x by the name of this sentence.

> (2) x is a true sentence if and only if p.[5]

In "The semantic conception of truth," Tarski puts it less succinctly:

> The sentence "snow is white" is true if, and only if, snow is white.[6]

This is Tarski's T-schema (also referred to as Tarski's biconditional, T-convention or material-adequacy condition), which can be used as a recipe to provide the truth conditions for any sentence.

If we name a sentence by putting quotation marks around the sentence itself, we get a formulation of the type "'p' is true iff p", where p can be replaced by any sentence. This is often how the T-schema is stated. With this formulation, the second 'p' is the sentence itself, and the first ''p'' is the name of that sentence. However, this restricts the schema to instances where the same language can be used to express both the sentence itself and the name of the sentence. Tarski therefore prefers using two different symbols (x and p) in order to clearly differentiate between the sentence itself and its name. Thus x is the name of the sentence p.

[4] Ibid., 155
[5] Ibid.
[6] Tarski, "The semantic conception of truth," 334.

The T-schema is meant to be a structure that provides truth conditions for any sentence, by replacing p by the sentence in question. Thus, if we replace 'p' by the sentence 'the sea is blue', the T-schema gives us "'The sea is blue' is true iff the sea is blue."

Any name, such as 'x', 'S_1' or 'Æthelflaed' can be given to a sentence. In this instance the T-schema reads "S_1 is true if and only if the sea is blue."

Sentences (or at least what they express) can also be referred to by sentences in other languages. The T-schema "'The sea is blue' is true iff the sea is blue" can also be expressed by

"Mare caeruleum est" is true if and only if the sea is blue;

or by

"海是蓝的" is true if and only if the sea is blue.

The T-schema's succinct "'p' is true iff p" formulation is insufficient to accommodate references to the sentence that are not the sentence itself and explains why Tarski uses the formulation "x is a true sentence if and only if p" where 'x' can be replaced by the name of a sentence and 'p' by that sentence.

The problem with the T-schema is that it falls prey to the liar paradox. When we take the sentence 'The sentence λ is false', and give this sentence the name λ, the liar paradox emerges.

Tarski thus concludes that such a definition of truth "meets with very real difficulties."[7]

2.1.2 Second attempt: the structural definition

Tarski then attempts a second type of definition, which he calls "structural". This is an attempt at picking out true sentences based on their logical form. For instance, sentences of the form "If A then A", where A is any sentence, are always true. In this way, composite truths can be found through deduction rules governing the use of logical operators.

However, this attempt quickly fails as well. Natural language does not have a set of clearly defined rules governing which sentences belong to the language and how to build new ones. In Tarski's words, natural language is not "finished, closed, or bounded by clear limits."[8] Establishing a truth concept based entirely on a sentence's syntactical form is ill-fated.

[7] Tarski, "The concept of truth in formalized languages", 162.
[8] Ibid., 164.

Natural languages do have syntactical grammatical rules aiming at describing how sentences can be built out of words, but these rules are never more than an attempt at describing a language at a particular time. Languages are in constant change and speakers of the language are always inventing new syntactical variants that sometimes move from an individual's own idiolect to a standard form in either a regional form of the language or a sociolect, or into the language itself. The noun "chair", for example, was in use in English for centuries before it also became a verb. A sentence such as "She is chairing the meeting" would have been deemed syntactically incorrect in the 19th century, but is attested from 1921[9] and is now considered correct.

Furthermore, even if it were possible to set up an exhaustive set of syntactic rules for the elaboration of sentences, semantics would still have to be taken into account. Many sentences are grammatically correct but are seen as nonsensical, and there is a lot of debate about whether or not they are part of the language. One example is Chomsky's famous phrase "Colorless green ideas sleep furiously".[10] Chomsky uses it as an example of a sentence that is syntactically well-formed yet semantically nonsensical, since something cannot be both colourless and green at the same time, ideas neither have a colour nor sleep, and the activity of sleeping cannot be done furiously. Nevertheless, whether or not Chomsky's sentence is meaningless, meaningful but nonsensical, or indeed meaningful and intelligible, as several poetic interpretations have claimed,[11] remains an open question, which shows how difficult and contentious this topic is. Another example is Moore's paradox: "It is raining, but I believe that it is not raining."[12] Despite being syntactically correct, this is a very odd sentence, and there remains a sense that this is simply not how language works.

Syntax alone is not sufficient for determining which sentences are part of a language, let alone for deciding which sentences are true. Elaborating a structural syntactic definition of truth that would provide us with rules to check whether a sentence is true seems therefore to be doomed to fail. Tarski concludes: "The attempt to set up a structural definition of the term 'true sentence'—applicable to colloquial language is confronted with insuperable difficulties."[13]

2.1.3 The impossibility of defining a truth predicate

[9] OED, "chair, v.".
[10] Chomsky, *Syntactic structures*.
[11] Erard, "The Life and Times of 'Colorless Green Ideas Sleep Furiously'"
[12] Moore, "Moore's Paradox".
[13] Tarski, "The concept of truth in formalized languages", 164.

After the failed semantic and structural attempts, Tarski shows more generally that no natural language can have a consistent true concept. He picks out three features of natural language that taken together lead to the liar paradox, thus showing that a natural language cannot have a consistent truth predicate. He writes:

> No consistent language can exist for which the usual laws of logic hold and which at the same time satisfies the following conditions:
>
> (I) for any sentence which occurs in the language a definite name of this sentence also belongs to the language;
>
> (II) every expression formed from (2) ["x is a true sentence if and only if p"] by replacing the symbol 'p' by any sentence of the language and the symbol 'x' by a name of this sentence is to be regarded as a true sentence of this language;
>
> (III) in the language in question an empirically established premiss having the same meaning as (α) [i.e. the sentence which asserts that the denoting term which occurs in the liar sentence refers to the sentence itself][14] can be formulated and accepted as a true sentence.[15]

The first condition (I) establishes the existence of self-reference in natural language. It states that the language includes the names of all its sentences, that is, that the language can refer to its own sentences. This rests on an important presupposition Tarski makes, namely that language is universal, and that everything can be translated into it and stated in it. Therefore, natural language must also accept expressions such as "'true sentence', 'name', 'denote'"[16] and names of sentences are part of the language.

The second condition (II) establishes the T-schema. Any sentence that follows the T-schema ("x is a true sentence if and only if p") is true. This is a condition on the truth predicate itself.

Finally, the third condition (III) simply states that the language must be capable of expressing that the liar sentence uses its own name. That is, if the liar sentence is stated as 'λ is false', then the language must be able to express that the sentence 'λ is false' is itself named 'λ'.

[14] Square brackets from Maudlin, *Truth and Paradox: Solving the Riddles*, 17.
[15] Tarski, "The concept of truth in formalized languages", 165.
[16] Ibid., 164.

Tarski thus establishes that any language which admits self-reference, considers sentences following the T-schema as true, and is able to express the liar sentence cannot have a consistent truth concept. Natural language satisfies all three conditions, and according to Tarski it is therefore impossible to avoid the liar paradox and define a truth predicate in natural language. In the rest of the paper, he elaborates a formal language without self-reference which admits a truth concept satisfying the T-schema. (See chapter 2 section 3 for more details about his solution).

That a truth concept cannot be precisely defined in natural language is not necessarily an issue for Tarski, as on the one hand everyone has "an intuitive knowledge of the concept of truth"[17] and on the other hand truth in natural language is discussed in epistemology.

2.2 Tarski's unspoken premise

Several times, Tarski appeals to "elementary logical laws"[18] or "the usual laws of logic"[19]. This refers to the laws of classical logic, which Tim Maudlin[20] reformulates as a fourth condition. Tarski's conditions can thus be re-stated:

(I) Self-reference can be expressed.
(II) Sentences following the T-schema are considered true.
(III) The liar sentence can be expressed.
(IV) Classical logic holds.

Two of these conditions, (I) and (III), can be combined: what we need is a language that is capable both of self-reference and of formulating the liar sentence. In other words, we need a language whose sentences can refer to their own truth or falsity, such as natural language.

We can therefore give a new recipe for the liar paradox:

(1) A truth concept which follows the T-schema.
(2) A language whose sentences can express their own truth-value.
(3) Classical logic.

[17] Ibid., 153
[18] Ibid., 162
[19] Ibid., 165
[20] Maudlin, *Truth and Paradox: Solving the Riddles*, 17.

Tarski took classical logic for granted and did not explicitly make it into its own condition. However, doing so opens another possibility: instead of giving up (1) or (2), why not consider giving up (3) and looking at non-classical logics? In other words, instead of either only considering formal languages without self-reference and unable to formulate the liar sentence, or adopting a truth predicate that does not satisfy the T-schema, it is also possible to remain in natural language (which can express both self-reference and the liar sentence) and keeping the T-schema by giving up classical logic.

This thesis will therefore choose to keep (1) and (2) and explore non-classical logics that avoid the liar paradox. The next section will provide some justification for this choice and chapter 2 will defend self-reference beyond its presence in natural language. Since it is well-established that natural language can express the liar sentence (see section 1 of this chapter), Tarski's third condition (III) will not be further addressed. Chapter 3 will defend the T-schema and chapters 4, 5, 6 and 7 will focus on non-classical solutions to the liar paradox.

Section 3 – Methodology and framework

3.1 Philosophy: between descriptivism and prescriptivism

Any philosophical inquiry into an abstract concept, such as truth, faces a question about its goal. Is the role of philosophy to describe the world as it is, and help us achieve better clarity in what concepts mean? Or is philosophy a prescriptive endeavour, aiming at changing our concepts to what they ought to be? Does the philosopher describe a pre-existing truth concept, or do they decree what this concept ought to be and how we should understand it?

Whereas the field of linguistics, which has also struggled with this question, has now come firmly out on the side of descriptivism, philosophy has not achieved any such consensus. Nor, perhaps, should it. Philosophy studies concepts that do not necessarily have an independent ontological status. Do the concepts of justice, freedom and beauty exist on their own, to be discovered and described? Or do these concepts in a sense not really exist, letting philosophers be free to shape and mould them as they please?

Extreme versions of both descriptivism and prescriptivism can be problematic and commitment-heavy. A strong descriptive option might assert that there exists a well-defined truth concept with an independent ontological status and that the philosopher's role is to discover it. Without launching into a deep discussion about the ontological nature of abstract entities and social kinds, there can be something quite odd about picturing the concepts of truth

or justice as some kinds of immaterial shimmering entities, waiting 'out there in the world' for the philosopher to find and to describe. This is nevertheless the view defended by some Platonists such as Frege,[21] who defends the existence of a realm of abstract objects including truth, and by some Christian theologians.[22] However, such views build on rather strong ontological presuppositions and commitments, and can therefore be difficult to adopt.

There are descriptive alternatives that make weaker ontological claims, for example by arguing that concepts are constructed by our usage of them. The role of philosophy would then be to describe how the concept is used by speakers, that is, to describe linguistic practices surrounding the concept. This does not require the existence of an underlying natural kind, as Haslanger stresses: an inquiry into justice, for example, may start with examining instances considered as just or unjust, in order to "determine whether there is an underlying (possibly social) kind that explains the temptation to group the cases together."[23]

Although this can be an important enterprise in itself, speakers and linguistic communities can be confused about a concept, or plain wrong. What is considered "just" or "beautiful" is subject to change and culturally dependent: how speakers in one community use the word 'just' does not settle once and for all the concept of justice. On the contrary, one could argue that although justice and beauty are important objects of study for philosophers, the goal is not to conclusively define and describe what they are, and to close any further enquiry. A great number of people are daily engaged in activities aiming at a more just society, and pursue aesthetic goals in their everyday life, and are constantly exploring and shaping the limits of the concepts.

The extreme version of prescriptivism is equally as problematic, however. If philosophy was entirely free to shape concepts and decide what they should mean, a philosopher could redefine a concept in a way so removed from its ordinary meaning as to become completely unrecognisable. This is what conceptual engineers such as Cappelen call a "change of topic"[24]. If I were to introduce a beautiful and coherent new concept, call it 'truth', even though it was very far removed from normal linguistic practices, and claim that we now had a perfectly working truth concept avoiding all semantic paradoxes, I am quite certain no-one would accept my claim and adopt this new concept. It would not solve any issues linked to the old concept, and would not be of any use.

[21] Frege, "Der Gedanke. Eine Logische Untersuchung", translated as "Thoughts".
[22] Riga, "On Truth: A Catholic Perspective."
[23] Haslanger, "Gender and Race: (What) Are They? (What) Do We Want Them to Be?", 33.
[24] Cappelen, *Fixing Language: An Essay on Conceptual Engineering*.

In some cases, it is argued that old concepts should be eliminated entirely and replaced by new ones, which may indeed be different enough so as to become unrecognisable. Kwame Anthony Appiah, for instance, argues that there is nothing in the world that the concept of race refers to, and that the concept should therefore be eliminated.[25] Better concepts, such as ethnicity, as Haslanger[26] proposes, should be used instead. When it comes to the concept of truth, however, we do not have the same ethical concerns to deal with as with the concept of race. Some philosophers, such as Kevin Scharp,[27] do indeed argue that the inconsistency revealed by the liar paradox makes the truth concept so deeply flawed as to be unsalvageable, and argue for it to be eliminated altogether, and replaced by two successor concepts. But Kevin Scharp's successor concepts are still based on the old concept of truth, and do not change the topic entirely. The truth concept, in its ordinary daily usage, is not only fairly unproblematic but also ubiquitous and even necessary, for instance in legal contexts where lies can result in serious criminal charges. It cannot be eliminated or replaced by an unrecognisable replacement concept.

Philosophy has to contend with the tension between descriptivism and prescriptivism. We do not want to be prisoners of antiquated and faulty concepts, but neither can we change them to something completely unrecognisable. Any inquiry into a concept must be sufficiently grounded in linguistic practice so as to avoid changing the topic entirely, yet some measure of freedom must be maintained to be able to present solutions to paradoxes or other issues.

3.2 Conceptual engineering

Some philosophers argue that their role is to be "conceptual engineers". Many philosophical puzzles and problems cannot be solved by elaborating a better theory; the concepts at play themselves are faulty. Although conceptual engineering is explicitly normative and prescriptive, that is, they mean to tell us what concepts we ought to have, it still needs to avoid changing the topic entirely, and thus also has to deal with the tension between descriptivism and prescriptivism. Cappelen[28] speaks of "revision" when an old concept is improved (Chalmers calls it "conceptual re-engineering"[29]) and "replacement" when a new concept is introduced to replace the old ("*de novo* conceptual engineering" for Chalmers). It is not easy to

[25] Appiah, "Race, Culture, Identity: Misunderstood Connections."
[26] Haslanger. "Gender and Race: (What) Are They? (What) Do We Want Them to Be?"
[27] Scharp, *Replacing Truth*.
[28] Cappelen, *Fixing Language: An Essay on Conceptual Engineering*.
[29] Chalmers, David. "What is conceptual engineering and what should it be?"

know where the line goes between replacement and revision, especially since a replacement concept may keep the same name as the old concept.

Robyn Dembroff[30] acknowledges that their analysis of sexual orientation can be seen both as revising the old concepts or replacing them, but whether it is one or the other does not matter: "This tension is fine; I'm not sure anything important hangs on whether my project is described as providing a revised or replacement concept of sexual orientation"[31]. What matters is that their definition of sexual orientation must preserve enough features of the old concepts so as to be perceived to be about the same thing, yet provide sufficient changes to favour a social and political goal, which is having a theory of sex and gender that helps fight discrimination and promotes equality.

In this, Dembroff explicitly follows Sally Haslanger's 'ameliorative project', which is part of a different typology of philosophical methods. In *Gender and race*[32], Haslanger puts forth three types of strategies. The first is concerned about the meaning of the word or concept. Here the focus is on the word itself and its ordinary use among speakers. The second type of strategy, on the other hand, is concerned with the concept's extension in the world, and not merely with the content of the word. Dembroff phrases this as "ask[ing] which natural kind (if any) our ordinary concept of x tracks,"[33] and can also be an inquiry into social kinds.

The last type of strategy, also embraced by Dembroff, is the one Haslanger defends for her work on race and gender, and which she alternatively calls 'ameliorative project' or 'analytical project'. It is concerned with the role a concept plays within a specific theory. What is the concept's job description according to a specific theory? Which conceptual understanding is the most useful? This is what leads Haslanger to defend a definition of 'woman' based entirely on gender hierarchy, as this is for her the definition of 'woman' that is most useful for achieving the political goal of gender equality.

Haslanger's three strategies cannot be neatly classified as descriptive or prescriptive. At first glance, the two first strategies seem descriptive, since they are concerned with describing an ordinary concept or its extension in the world. However, such projects may also propose some changes or improvements of the concept or its extension (especially when the concept does not 'carve at the joints' of a clearly delineated natural kind), and thus become prescriptive. The

[30] Dembroff, "What is Sexual Orientation?"
[31] Ibid., 4.
[32] Haslanger. "Gender and Race: (What) Are They? (What) Do We Want Them to Be?"
[33] Dembroff, "What is Sexual Orientation?", 3.

third ameliorative strategy, even though it proposes a change of the concept and prescribes how we ought to reconceptualise our understanding of certain concepts, is nevertheless anchored in the description of ordinary concepts, in order to avoid the pitfall of changing the topic entirely.

Haslanger's ameliorative project is not well suited for the study of truth. The ameliorative project puts ethics first: the best concept is the one that fits into the theory with the most ethical social and political goals. Although this makes sense in the context of concepts such as sex, gender, race, disability and so on, it is not at all obvious what kind of theory involving truth would be the most ethical. According to what criteria should we rank theories involving truth? The only theory requirement to follow is to have a somewhat usable truth concept in our conceptual toolbox, which means that the liar sentence must be dealt with somehow. When it comes to truth, choosing a theory first, which is then used to decide the meaning and extension of the concept, is not a great idea. It would be possible to elaborate a general theory involving semantic concepts other than truth first, and only thereafter choose a truth concept that fits into this theory. However, there is no reason why other semantic concepts should have primacy over truth and why they should affect its definition. There's just as good a reason to choose a truth concept first and have it guide the way we think of other semantic concepts.

3.3 Strategy for this thesis

This thesis will also have to deal with the tension between being descriptive (explicating what truth is) and prescriptive (establishing what truth should be). However, it will attempt to be as neutral and open-minded as possible towards pre-existing theories.

The two guiding criteria will be the following: getting a truth concept that is recognisable as the everyday truth concept, while dealing with the inconsistency revealed by the liar paradox.

In order to remain close to the everyday truth concept and avoiding changing the topic, the thesis will investigate truth in the context of natural language, and not within that of a formal artificial language. Furthermore, it will take Tarski's T-schema as a requirement, based on empirical considerations (empirical linguistics and surveys among philosophers and non-philosophers, see chapter 3).

The inconsistency revealed by the liar paradox will be dealt with by giving up on classical logic. This is not as radical as it might seem at first: the liar sentence is only problematic when confronted with two specific logical principles, namely the principle of explosion, which states that any proposition, no matter how false or nonsensical, can be derived from a contradiction,

and the principle of excluded middle, according to which every proposition must be either true or false. When these two principles are followed, every proposition, liar sentence included, must be assigned one truth-value, true or false, and one truth-value only.

These two principles do not hold in certain non-classical logical systems. In order to keep natural language and the T-schema, we will therefore have to move the truth concept into a logical system without these two principles, either into a logic that can accept that a third truth-value is assigned to the liar sentence, or into one where inconsistency is made palatable. The logical system that this thesis will ultimately defend, First Degree Entailment (FDE), is still closely related enough to classical logic so as to avoid being accused of "changing the topic" and defending an unrecognisable logical system (see chapter 6). The concept of truth in FDE thus manages to handle the liar paradox while upholding the T-schema in the context of natural language.

Chapter 2 – In defence of self-reference

Self-reference is one of the causes of the liar paradox, and is the one Tarski decides to give up on. It may therefore easily be seen as the weakest link, which ought to be attacked first, especially when the alternative is to give up classical logic. However, self-reference is a natural and generally unproblematic feature of ordinary language (section 1). Furthermore, self-reference emerges very easily in arithmetic (section 2), even though it is a very simple system compared to natural language. Avoiding self-reference would therefore be impractical and cumbersome. The last section (section 3) will present Tarski's own solution without self-reference, and will argue that the concept of truth within such a theory is too far-removed from our ordinary understanding of truth.

Section 1 – Self-reference in natural language

This sentence is written in English. The previous sentence says something about itself and is utterly unproblematic. It is difficult to imagine how natural language could be restricted so as to ban all self-referential sentences. Even if such a thing was possible, and syntactical rules were revised to make such sentences ungrammatical, self-reference has a tendency to emerge accidentally and would cause the liar sentence to rear its head.

Imagine an art historian walking past the classroom of an older colleague, classroom 1593, and seeing on the wall the sentence "There were no important female Baroque painters." Furious, and thinking about the celebrated seventeen-century painter Artemisia Gentileschi, our younger art historian goes to her own office and writes "The sentence in room 1593 is false" on a blackboard. Later, the sentence in classroom 1593 is erased and the blackboard is itself moved to this classroom. Suddenly, the sentence no longer means that there were in fact important female Baroque painters, but has inadvertently become a liar sentence. Thus, even restricting language so as to avoid sentences referring to themselves would not suffice to avoid self-reference.

It is true that it is not self-reference itself that is a problem: 'This sentence has five words' is self-referential yet not problematic. The issue is having sentences that say something about their own veracity, in particular when asserting their falsity. 'This sentence is true' is not as problematic as the liar sentence, but is nonetheless puzzling, as no truth-value can easily be assigned to it. However, banning sentences from referring to their own truth or falsity would

not suffice: as the story from the art history department tells us, one would have to keep sentences from expressing truth and falsity altogether. (This is Tarski's solution, developed in section 3 of this chapter.)

Natural language, however, does not bow to externally imposed rules. Languages are continuously changing, evolving and adapting so as to express new ideas and nuances. Most of the time, these changes occur organically: individuals and groups constantly make alterations and adjustments; some of these are interpreted as linguistic mistakes, some are used in a limited social group for a limited amount of time, while others spread and end up changing the language itself. If there is a need to express "The sentence room 1593 is false", any language would come up with a way to express this.

Tarski himself, in "The concept of truth in formalized languages" makes an important presupposition that supports this point. Tarski claims that language is universal and that anything can be translated into it.[34] If something can be expressed in one language, any other language can be adapted to express the same, even when this requires the introduction of some new vocabulary or other types of linguistic innovations.[35] Some nuances, especially poetic, can be lost in translation, and some languages are perhaps better equipped than others to express certain things. However, expressing both self-reference and a sentence's veracity are both quite straightforward; attempting to restrict natural language from expressing either would never last long.

All of this supports Tarski making the presupposition he does. Sentences in natural language have the ability to refer to and talk about the truth-value of any sentence, including themselves. Restricting the ability of natural language to avoid this would distort the nature of language and would prove impossible.

Section 2 – Self-reference in arithmetic

Self-reference emerges very easily in arithmetic. By a technique known as Gödel numbering, any statement about arithmetic can be mapped onto a number. This way, statements about

[34] Tarski, "The concept of truth in formalized languages".

[35] To what extent this is the case has been the subject of a lot of debate within linguistics (see Talmy, "Universals of semantics"). However, historical linguistics (see Millar, *Trask's Historical Linguistics*) shows how whenever a language has needed to express something new, it always finds a way to do so, even by merely taking a word or expression from another language.
See also Simmons (*Universality and the liar*) who argues that natural linguistics is semantically universal, that is, natural language can say anything about its own semantics.

arithmetic are no longer merely part of a 'meta-language' that can talk about arithmetic but which is not itself part of it. These statements become part of arithmetic itself, which thus gains the capacity to represent statements about itself. The diagonal lemma goes one step further and proves the existence of self-referential sentences.

2.1 Gödel numbering

In order to set up the Gödel numbering, we first need to set up an arithmetic formally. Let us consider Peano arithmetic, which is simply the natural numbers $\mathbb{N}$ with addition $+$ and multiplication $\times$. Formally, all that is needed for the natural numbers are 0 and a successor function called S. With these, we can define 1 as the successor of 0 and write it as $S0$. 2 is the successor of 1, so it is written $S1$ or $SS0$, and so on. We also add the identity $=$.

In order to express statements about arithmetic we need some simple logical operations: negation $\neg$, disjunction $\vee$ and implication $\rightarrow$. We will also use the existential quantifier $\exists$ and some syntactical symbols: parentheses and comma.

First, we assign a number from 1 to 12 to each of the twelve symbols needed to express an arithmetical statement. Variables (*x*, *y*, *z* etc.) map on prime numbers greater than 12. Every statement about arithmetic can now be expressed using these symbols and the variables.

Gödel numberings can be set up in many different ways; the one presented here is merely one possible method.

Logical symbols	Meaning	Gödel number
0	Zero	1
S	Successor function	2
$+$	Addition	3
$\times$	Multiplication	4
$=$	Identity	5
$\neg$	Negation	6
$\vee$	Disjunction	7
$\rightarrow$	Implication	8
$\exists$	Existential quantifier	9
(	Left parenthesis	10
)	Right parenthesis	11

,	Comma	12

Let us encode the statement $2 + 1 = 3$. Rewritten with the help of the successor function, it becomes $SS0 + S0 = SSS0$. The codes for these symbols are respectively 2, 2, 1, 3, 2, 1, 5, 2, 2, 2 and 1.

Next, we use prime numbers to encode the string, by raising each successive prime number to the power of the code of symbol, which gives us $2^2 \times 3^2 \times 5^1 \times 7^3 \times 11^2 \times 13^1 \times 17^5 \times 19^2 \times 23^2 \times 29^2 \times 31^1$. Since every number has a unique prime factorization, this encoding is unique. In this way, the statement $2 + 1 = 3$, which says something about arithmetic, can be expressed through one single number.

Even very simple statements map onto very big numbers, so this is obviously not a very efficient way of encoding statements. However, this is not meant to be used in practice, but simply to show that there is a way to map any statement about arithmetic onto a number. With this, arithmetic is able to represent statements about itself.

For any sentence or formula A, we write $\ulcorner A \urcorner$ for its Gödel number.

2.2 The diagonal lemma

The Gödel numbering technique is central to the diagonal lemma, which proves the existence of sentences that refer to themselves in certain formal theories such as the Peano arithmetic.

Diagonal lemma: Let T be a first-order theory in arithmetic. For any formula $\Phi(x)$ there is a sentence F such that $\vdash_T F \leftrightarrow \Phi(\ulcorner F \urcorner)$.

The proof will not be presented here, but can be found in most mathematical logic textbooks such as Boolos[36] or Mendelson[37].

The diagonal lemma says that for any formula with a free variable $\Phi(x)$, there exists a sentence that is equivalent to the formula applied to the Gödel number of that sentence. This means that there are sentences that are equivalent to a formula expressing something about that sentence. In other words, these sentences refer to themselves and self-reference emerges in arithmetic.

[36] Boolos, Burgess and Jeffrey, *Computability and Logic*, section 17.1, 221.
[37] Mendelson, *Introduction to Mathematical Logic*, section 3.34, 205.

Getting the liar sentence is now very easy. First, we need to introduce the truth predicate Tr such that $Tr(x) := 'x\ is\ true'$. Let $\Phi(x)$ be the formula $\neg Tr(x)$. This is the mathematical way of saying that $\Phi(x)$ expresses 'x is false'. According to the diagonal lemma, there exists a sentence λ such that $\lambda \leftrightarrow \Phi(\ulcorner \lambda \urcorner)$, that is, $\lambda \leftrightarrow \neg Tr(\ulcorner \lambda \urcorner)$. The sentence λ is now equivalent to 'λ is false'. In order words, the liar sentence can be expressed in a simple arithmetic.

This shows that self-reference and the liar sentence not only appear in natural language, but also in much simpler formal languages such as arithmetic.

Section 3 – Tarski's solution without self-reference

According to Tarski, the T-schema is an integral feature of truth, and classical logic is a given. He therefore chose to give up on self-reference in order to keep the T-schema and classical logic. More precisely, he defined a truth predicate based on the T-schema, but in a formal language in which sentences were not able to express anything about their own truth or falsity. Since the liar sentence says of itself that it is not true, it cannot be formulated in such a language.

A language without a truth predicate applying to its own sentences is called "semantically open": its truth predicate applies only to sentences in another language, and the language is not self-contained. A language whose truth-predicate applies to itself is on the contrary called "semantically closed". Natural language is such a semantically closed language.

Tarski establishes a distinction between the object language and the metalanguage. The object language is a language without truth predicate, that cannot express the truth or falsity of its own sentences. Only a language at a 'level above', a metalanguage, can express the truth or falsity of sentences in the object language. The metalanguage will always include the object language, but will also contain additional elements that allows it to express the truth or falsity of sentences in the object language. The metalanguage contains a truth predicate, but this truth predicate can only be applied to the object language. The truth predicate in the metalanguage cannot be used to express the truth or falsity of sentences in the metalanguage itself.

In order to do that, we need to consider the metalanguage as an object language and thus establish a 'meta-metalanguage', containing the metalanguage itself as well as a new 'meta-truth predicate' that can be applied to this metalanguage.

In short, we end up with a layering of languages, with the object language at the bottom, and successive metalanguages above. Each metalanguage contains a truth predicate that can only be used to express the truth or falsity of the language right below.

Tarski never meant for this solution to establish a truth predicate for natural language that would solve its liar paradox. For Tarski, natural language simply cannot have a consistent truth predicate. The object language/metalanguage construction is meant to be artificial, and truth within that context was not intended to be the same concept as our everyday understanding of it. Tarski establishes a working definition of truth in an artificial formal language, but if our goal is to search for a truth predicate in natural language, we cannot embrace his solution.

Chapter 3 – In defence of the T-schema

Whether the T-schema is right or not is still an ongoing debate, and it would be far beyond the scope of this chapter to comprehensively present all the arguments on each side. I will not assert that the T-schema is a necessary component of any theory of truth, but instead I will merely argue that the T-schema does seem like a common minimal requirement for truth. It is compatible with the two main truth theories, deflationism and correspondence theory, which are the two main theories held by professional philosophers. Indeed, more empirical data seems to suggest that the T-schema is at the very least consistent with an ordinary understanding of truth. The last section will present a few truth theories opposing the T-schema. Nevertheless, there is still enough evidence to justify preserving the T-schema and exploring what happens in non-classical logic.

Section 1 – The T-schema

The second partial cause of the liar paradox is having a truth concept that follows the T-schema. As we saw in chapter 1, Tarski introduces the T-schema as a biconditional between the truth of a sentence and the sentence itself. He writes, "x is a true sentence if and only if p",[38] where p is a sentence and x is the name of this sentence. Any formula that follows this schema, such as "The sentence 'Emmy Noether was a mathematician' is true if and only if Emmy Noether was a mathematician" is an instance of the T-schema.

In order to state the T-schema more formally, we need a truth predicate, which we will write as T. For the sake of simplicity, we will assume that every sentence has a name: for a sentence A we will use the notation $\langle A \rangle$ to denote its name[39]. In order to say that a sentence A is true, we will write $T\langle A \rangle$.

Tarski states his T-schema using the phrase "if and only if," and there are in fact different ways to formalise it. I will present three of them, the first of which uses a biconditional, the second a bi-entailment, and the third which considers the T-schema to be the combination of two axioms.

The most typical way to formalise the T-schema is by using a biconditional: $T\langle A \rangle \leftrightarrow A$. The T-schema is then considered to be a rule added to the language.

[38] Tarski, "The concept of truth in formalized languages", 155.
[39] Following the notation from Beall, Glanzberg and Ripley, *Formal theories of truth*.

This biconditional can be analysed as consisting of two parts, corresponding to each direction of the biconditional. From left to right, the T-schema says that if a sentence is true, then we can state the sentence: $T\langle A\rangle \rightarrow A$. This has been called "release"[40], as the truth predicate seems to release the sentence.

From right to left, the T-schema states $A \rightarrow T\langle A\rangle$, that is, if we have a sentence, then we can state that the sentence is true. This is also called "capture", as the truth predicate captures the sentence.

The second way to formalise the T-schema uses logical entailment. Indeed, according to *capture*, a sentence A entails $T\langle A\rangle$, which can be written $A \vdash T\langle A\rangle$. Conversely, according to *release*, $T\langle A\rangle$ entails A, which is written $T\langle A\rangle \vdash A$. In other words, the T-schema can also be stated with a bi-entailment, as $T\langle A\rangle \dashv\vdash A$.

Finally, the T-schema can also be understood as two separate axioms, one corresponding to *capture*, the other to *release*. Hartry Field calls them respectively "T-IN" and T-OUT."[41]

(T-IN) $\vdash$ If A then $T\langle A\rangle$

(T-OUT) $\vdash$ If $T\langle A\rangle$ then A

Tarski does not give much arguments in favour of the T-schema, and seems to consider that the T-schema captures the behaviour of the truth concept in a way that is intuitively quite obvious.

Section 2 - T-schema, correspondence theory and deflationism

Tarski has been interpreted both as a correspondence theorist (for instance by Field[42]) and as a deflationist (for instance by Soames[43]). Correspondence theory was dominant when Tarski wrote his papers, which is why it may have been natural to interpret him as a correspondence theorist; however, his theory can also be interpreted through a deflationist lens. This section will present these two possible interpretations.

2.1 Tarski and the correspondence theory

According to the correspondence theory of truth, truth consists in some sort of correspondence between a sentence and reality. A sentence is true if what it says is a fact. Truth is a relation

[40] This terminology is for example used by Beall, Glanzberg and Ripley, *Formal theories of truth*.
[41] Field, "Truth and the Unprovability of Consistency."
[42] Field, "Tarski's theory of truth."
[43] Soames, "What is a theory of truth?"

between a sentence and some portion of reality. It is more correct to speak about the family of correspondence theories, as there are multiple theories with different notions of what exactly this relation is ("correspondence, conformity, congruence, agreement, accordance, copying, picturing, signification, representation, reference, satisfaction"[44]) and which portion of reality the sentence relates to ("facts, states of affairs, conditions, situations, events, objects, sequences of objects, sets, properties, tropes"[45]).

Tarski's T-schema does indeed seem to fit quite well to the correspondence theory: "$'p'$ is true iff p" does indeed seem to refer to a correspondence between the sentence p and what the sentence p says about the world. A sentence is true if it says something about the world that is indeed true: there is some kind of relationship between the sentence and reality. If the sentence "Karen Uhlenbeck won the Abel prize in 2019" is true, it means that there is indeed such a person as Karen Uhlenbeck who won the Abel prize three years ago, in the real world. In the formulation $T\langle A\rangle \leftrightarrow A$, '$T\langle A\rangle$' refers to a sentence being true and 'A' is the sentence itself, which says something about the world. The biconditional thus establishes a relationship, a correspondence, between a sentence being true and something happening 'out there in the world.' Field[46] interprets Tarski as a correspondence theorist: a sentence is true if its parts refer to reality. This reference relation is a "physical or causal relation between words and the world"[47].

2.2 Tarski and deflationism

Tarski can also be interpreted as a deflationist, however. According to the deflationary theory of truth, there is nothing substantial about truth. This position has for instance been defended by Field[48] and Horwich[49]. According to deflationists, correspondence theorists are mistaken: there is no correspondence or relation to be found between a sentence and the world. Saying that a sentence is true is equivalent to stating that sentence. Stating that "it is true that Ada Lovelace was the first computer programmer" says nothing more substantial than simply stating "Ada Lovelace was the first computer programmer." There is nothing metaphysically interesting about the concept of truth.

[44] David, "The Correspondence Theory of Truth"
[45] Ibid.
[46] Field, "Tarski's theory of truth."
[47] Lynch, Michael P. "Realism and the Correspondence Theory: Introduction," 15.
[48] Field, Hartry. "Deflationist views of meaning and content."
[49] Horwich, Paul. "A defense of minimalism."

Tarski's T-schema can also be interpreted in this deflationary manner: after all, the T-schema establishes an equivalence between a sentence being true and the sentence itself, but says nothing about the nature of this equivalence, beyond the requirements of the biconditional. The T-schema $T\langle A\rangle \leftrightarrow A$ merely establishes an equivalence between 'A' and ''A' is true': stating $T\langle A\rangle$ is nothing more than stating A. Soames interprets Tarski in this way[50].

To conclude, Tarski's T-schema is compatible with both the correspondence theory and the deflationary theory of truth. It is impossible to say with certainty which position Tarski would have espoused, but for the purpose of this thesis that does not matter. What matters is that for Tarski, the T-schema is a necessary condition for any truth theory.

Section 3 – Some empirical arguments in favour of the T-schema

The T-schema can be interpreted both in a correspondence theoretical as well as in a deflationary way. In this section I will give some empirical evidence for the importance of these two theories. In contemporary philosophy in particular, there is a great deal of support for deflationism. There are many different versions of deflationism, which would be beyond the scope of this thesis to discuss. However, the vast majority of deflationsists defend the T-schema as central for the truth concept, and there is therefore significant support for it.

According to a 2020 survey, correspondence theory and deflationism are the two most common truth theories among philosophers. Furthermore, in a 1938 study, Arne Næss made a survey of how non-philosophers understand the concept of truth, and most answers are also compatible with the T-schema. The goal of this thesis is not to argue in favour of one theory of truth in particular; instead, I wish to remain as neutral and agnostic as possible, and merely remain close to a conception of truth that is recognised by ordinary speakers, despite the multiplicity of different truth theories. Although I do not arrive a decisive proof in favour of the T-schema, it provides some justification for considering that the T-schema is a minimal requirement for a truth theory close to the ordinary understanding of truth.

3.1 Truth-theories among philosophers

In 2020, David Bourget and David Chalmers[51] did a survey of English-speaking and English-publishing philosophers in order to assess what the current views of professional philosophers are on a range of philosophical topics. For the concept of truth, there were four options:

[50] Soames, "What is a theory of truth?"

[51] Bourget, David and Chalmers, David. *Philosophers on Philosophy: The 2020 PhilPapers Survey*.

"correspondence", "deflationary", "epistemic" and "other". A large majority answered in favour of correspondence theory or deflationism: 75.9% (with the inclusive method, 70.5% with the exclusive method). More precisely, 51.4% defended correspondence theory (48.3% excl.) and 24.5% (22.2% excl.) defended deflationism.

Since the T-schema can be interpreted both as a correspondence theory and as a deflationism, the results of the survey tell us that most professional philosophers have a conception of truth that is compatible with the T-schema.

3.2 Truth theories among non-philosophers

In 1938, Arne Næss published a monograph titled *"Truth" as conceived by those who are not professional philosophers*,[52] which is an empirical analysis of how non-philosophers talk about the abstract concept of truth. He notes in the introduction that among philosophers, the discussion about truth has already lasted 2500 years, has probably involved about a thousand people and "the number of standpoints felt as different or incompatible may be said to be 2, 100 or 1000"[53]. Tarski, who refers to Næss' study in the paper where he introduces his own truth theory, writes about "those endless, often violent discussions on [the right conception of truth]."[54] Indeed, the right conception of truth is still a contentious topic among philosophers today.

To add to his frustration about the in-fighting within philosophy on truth, Næss disagrees with the dismissive attitude many philosophers have to non-philosophers' opinion about the notion of truth. Many philosophers make claims about what "the man in the street"[55] thinks about truth, about what kind of folk-notion ordinary people have. But how can they know what non-philosophers actually think of the truth concept? Instead of being dismissive, perhaps some answers may be found in the understanding of truth 'ordinary people' have, since the polemical disagreements among philosophers are not being very useful. What Næss found is that instead of non-philosophers having similar simple and thoughtless views on truth, there was instead a great deal of diversity and complexity.

It has been claimed that Næss undertakes this study to attack the idea that correspondence theory, which was the dominant truth theory at the time, is intuitively obvious and that it

[52] Næss, *"Truth" as Conceived by Those Who Are Not Professional Philosophers*.
[53] Næss, *"Truth" as Conceived by Those Who Are Not Professional Philosophers*, 14.
[54] Tarski, "The semantic conception of truth, 348.
[55] Næss, *"Truth" as Conceived by Those Who Are Not Professional Philosophers*, 14.

coincides with the folk-understanding of truth. Carnap and Popper both thought that Næss' study undermined Tarski's theory, but this is very much disputed. Ulatowski offers an interesting discussion on this subject.[56] Næss does indeed criticise the idea that correspondence theory is the one that the person 'in the street' espouses, but as we have seen, Tarski's theory can be interpreted in different ways

My goal in this section is not to discuss whether or not Næss' study undermines Tarski's truth theory. Instead, I wish to argue that many 'intuitive' conceptions of truth by non-professional philosophers are compatible with the T-schema.

Næss' study can be difficult to analyse in terms of modern truth theories. In fact, Næss' purpose was not to group the responses according to pre-existing philosophical truth theories, but to give an overview of the incredible diversity of ideas present in non-philosophers. Næss made 37 different groups[57] based on the answers to his survey, and there is not much point in listing every single one of them.

Many groups are compatible with the T-schema: the most obvious ones are groups 1 to 4, which are variations around truth being a relation or an agreement, either with reality, real things or facts, and groups 5 to 8, which identify truth with something that happens, exists, is a fact or is the case. Other groups (11, 14, 15, 19, 28, 33) identify truth with what must be, what is evident and cannot be doubted or disproved, and can also be seen as compatible with the T-schema. There is a great deal of diversity: groups 13, 17, 18, 20 and 21 identify truth with empirical evidence, experience and senses; group 34 is some kind of a moral truth theory of what ought to be; other groups (22,23,24, 29, 30, 31) seem to espouse an epistemic conception relying on what one knows, has been taught or learned by testimony; and one (32) defends a pragmatic view where what is true is what is good to mankind.

Not all of these groups are compatible with the T-schema. However, one must remember that Næss' questionnaire was deliberately very open: he did not in advance distinguish between the many meanings 'truth' can have. Here, we care about truth as it is used in contexts such as "it is true that $2 + 2 = 4$", or "'Munch was a painter' is true". Næss, on the other hand, did not wish to exclude other senses of 'true', for instance as in "being a true friend", where 'true' is closer to 'loyal', or as in the New Testament, when Jesus says "I am the way, the truth and the life" (John 14:6), where truth seems to be some kind of metaphysical entity. Næss'

[56] Ulatowski, "Ordinary truth in Tarski and Næss."
[57] Næss, *"Truth" as Conceived by Those Who Are Not Professional Philosophers*, 66-69.

questionnaire started with open questions, of the type "What kind of things are true?" and "What are the common characteristics of true things?" As a result, not all answers are applicable for truth in the context of true sentences. Nevertheless, many of the groups are explicitly compatible with the T-schema.

Næss did in fact establish questionnaires to investigate non-philosophers' attitude towards the T-schema itself: whether they were willing "to substitute "p" for "p is true"[58]. However, very annoyingly for my purposes, he does not give the results and writes that "the results cannot be stated in a few words and must be omitted in this work"[59].

However, Tarski claimed that:

> in a group of people who were questioned only 15% agreed that "true" means for them "agreeing with reality," while 90% agreed that a sentence such as "it is snowing" is true, if and only if, it is snowing.[60]

The 15% is quite close to what Næss found, so I cannot help but wonder whether the second result might not also come from Næss, since Tarski and Næss were in contact. If it is indeed true that 90% agreed with an instance of the T-schema, despite not explicitly espousing a truth theory built upon such a schema, then this shows that the T-schema is indeed compatible with a range of views on truth, and that a truth theory, if it wishes to be compatible with intuitive understandings of truth, ought to be consistent with the T-schema.

3.3 Contemporary advances

It was surprisingly difficult to find evidence in favour or against the T-schema based on how truth is used in natural language. The ordinary conception of truth has not been a subject of much interest for philosophers, Tarski and Næss excluded. Comparative linguists such as Anna Wierzbicka who have studied truth across languages have not done it with the express purpose of testing how the T-schema holds up against truth in natural language across the world.

There is currently a new research project headed by Joseph Ulatowski, which is named *Truth without borders*[61] and which aims at continuing Næss' project and study intuitions

[58] Næss, *"Truth" as Conceived by Those Who Are Not Professional Philosophers*, footnote p.148.
[59] Ibid.
[60] Tarski, "The semantic conception of truth," 354-355
[61] https://www.josephulatowski.net/twb-truth

about truth among ordinary speakers. It is however far more ambitious in that it intends to examine truth concepts in various languages, in order to hopefully get a less Western-centric understanding of what we mean by truth. It will be very interesting to see the results, and especially whether some further support in favour of the T-schema can be found.

Although I cannot assert with certainty that the T-schema is a component of the everyday conception of truth, there is nevertheless evidence that supports taking the T-schema as a minimal requirement. Even while remaining as neutral as possible about what a full definition or theory of truth would look like, there is reason to think that the T-schema may very well be part of such a theory. At the very least, it is worth investigating what happens if the T-schema is kept and to explore non-classical options.

Section 4 – Against the T-schema

In the remainder of this chapter, I want to mention a few theories that oppose the T-schema. I will briefly consider coherentism, and Riki Heck's argument against disquotationalism, before addressing Kevin Scharp's attack on the T-schema

4.1 Coherentism

There are theories of truth that have been influential but which do not mention the T-schema, for instance coherentism.

According to a coherentist theory of truth[62], a proposition is true if it is in coherence with some other set of propositions already considered to be true. If there is a contradiction between a new proposition and some previously accepted proposition, the new proposition will be deemed to be false. The truth-value of a proposition is not dependent on how the world is. Although coherentism may seem odd at first, it describes quite well how agents actually decide whether or not to believe whether a statement is true. If a statement fits with what I already believe to be true, if it fits with my worldview, I am very likely to add it to my list of beliefs. If, one the other hand, a proposition goes against my understanding of the world and would require me to give up on too many other beliefs, I will probably be very suspicious against it and believe it to be false. For instance, if the sentence "Napoleon Bonaparte was born in 1269" was true, I would have to change my entire understanding of European history. I assume that such a sentence is

[62] Young, "The Coherence Theory of Truth"

wrong based on how it contradicts my previous knowledge, not by requesting to see the baptismal records of Ajaccio.

Even though the coherence theory of truth may describe quite well how sentences considered to be true are actually chosen, this mechanism may go horribly wrong. Some individuals, growing up in certain reclusive communities, for instance, may have entire belief systems that are wrong, and not in accordance to reality. There is of course an argument to be made that there is no objective reality, and that as long as a proposition coheres with someone's understanding of reality then it ought to be considered true. However, I believe such a view would take us too far from truth as it is commonly understood. Furthermore, coherentism is not necessarily against the T-schema, and is better considered as being neutral and not committed one way or another for or against the T-schema.

4.2 Heck's argument against disquotationalism

Riki Heck[63] argues against the T-schema in the context of a type of deflationism called disquotationalism. According to disquotationalism, not only is there nothing more to truth than the T-schema, like for deflationism in general, but the truth predicate in the T-schema is nothing but a device for 'disquotation.' Recall the T-schema "$'p'$ is true iff p." Truth is merely the device that allows us to go from a sentence in quotation marks, $'p'$, to the what the sentence says, p. I will give a very brief overview of their argument here.

According to Heck, the T-schema only works for some unproblematic sentences, such as "Snow is white" or "Susan Stebbing was a philosopher." These sentences stand on their own, and there is no particular obstacle to understanding them. However, many, if not most sentences, are not of this type. Context-dependence, indexicals and demonstratives are all features of sentences that can cause difficulties. Take the sentence "Mine is over there and it's fairly big." Saying that this sentence is true if and only if mine is over there and it's fairly big tells us nothing. Who is speaking? Whose thing are we speaking about? Where is 'over there'? What thing is it? What size is 'fairly big'? Are we talking about a fairly big pumpkin, a car or an apartment building? Merely removing the quotation marks tells us nothing at all about what it means for this sentence to be true.

However, Heck's attack is articulated as an attack against disquotationalism being a good theory of truth and context more than a direct attack against the T-schema itself. It is possible to agree

[63] Heck, "Truth and disquotation."

with Heck in that disquotationalists, and even perhaps deflationists, are missing out on something crucial about truth. However, that does not mean that the T-schema cannot play a role within a larger truth theory. Indeed, I can wish to keep the T-schema while also admitting that it simply pushes away the central issue of how a sentence relates to the world. I can paraphrase the sentence "Mine is over there and it's fairly big" being true as expressing that the object which the speaker refers to as hers is in a position appropriate to be called 'over there' for the speaker and has a size considered rather large in the context of whatever the object is. Of course, this sentence is extremely clunky, and there is still a lot of missing information that needs to be provided from context in order for it to gain meaning. Nevertheless, we need not abandon the entire concept of the T-schema.

As Heck's attack is formulated explicitly against disquotationalism, I will not discuss it forther here. Although I think that there is sufficient justification for exploring what happens to truth in non-classical logic if the T-schema is kept, which I do in the next chapters, it is important to note that there are strong voices against the T-schema.

4.3 Scharp's attack on the T-schema

Kevin Scharp's truth theory[64], on the other hand, is an explicit attack on the biconditional in the T-schema. According to him, the truth concept is irremediably broken and inconsistent, and he makes the choice of keeping self-reference and classical logic, and to attack the T-schema instead.

Scharp's solution can be seen as a 'splitting' of the T-schema into two conditionals: $T\langle p\rangle \rightarrow p$ (what we called 'release') and $p \rightarrow T\langle p\rangle$ ('capture'). Scharp proposes two new concepts that will replace the traditional truth concept: the first is called "descending truth" and corresponds roughly to the 'release' behaviour of T; the second is "ascending truth" and corresponds roughly to the 'capture' behaviour.

Since there is no longer one single concept that can both capture and release, the liar paradox cannot be derived. In fact, the liar sentence itself cannot be stated as it is, but gets replaced by two new liar sentences: 'this sentence is not descendent' and 'this sentence is not ascendent'. Without the biconditional, no paradox emerges.

Even though this solution can seem quite radical, as it amounts to saying that the old truth concept is wrong, flawed and needs to be thrown out in order to make space for two brand new

[64] Scharp, *Replacing truth*

concepts, it is not quite as drastic as it may look at first. In most cases, these two new concepts, ascending truth and descending truth, will overlap. When I write "It is true that Oksana Zabuzhko is a Ukrainian writer," it is both ascendent true and descendent true. In most contexts, the old concept 'true' can be seen as a shorthand for the two new ones. Semantic paradoxes occur at the very edge, where the two new concepts fail to overlap perfectly. Splitting the T-schema's biconditional will therefore only matter for semantic paradoxes which can now be avoided.

An argument against replacing truth with two new concepts can be found in a linguistics project called Natural Semantic Metalanguage, developed by Anna Wierzbicka and Cliff Goddard. The aim of the project is to establish universal concepts, called 'semantic primitives' or 'primes' found across all languages. The list is quite short (it contained 65 primes in 2013[65]), although it is constantly being expanded, as more linguistic research is done. This shows how very diverse human languages are, and that many words and concepts from our own languages we probably think of as universal may not be so at all. However, "true" appears on the list of primes[66], grouped together with other speech words such as "say" and "words."

The concept of truth is therefore a very basic and fundamental concept that every single one of the languages they have surveyed (across many language families) contains. Claiming that the truth concept should be abandoned and replaced by two new concepts thus appears quite suspect.

Of course, some traditional beliefs are utterly wrong and science has rightly replaced many of them. Perhaps the concept of truth represents a faulty belief, among the lines of Pythagoreans believing that fava beans contained the souls of the dead, and needs to be discarded. However, as we saw in chapter 1, it is rather controversial that the concept of truth should have an independent ontological status separate from human belief. If we replace the truth concepts by two new ones, will they still be recognised as 'true'? The inclusion of 'true' in the short list of universal primes makes it rather dubious, in my opinion, that it can ever be replaced by ascendent truth and descendent truth.

[65] Goddard and Wierzbicka, *Words and Meanings: Lexical Semantics Across Domains, Languages, and Cultures*, 12.
[66] Ibid.

It must be granted that this is not necessarily the purpose of Scharp's project. The two replacement concepts provide a technical solution to the liar paradox and other semantic paradoxes, and for everyday speech there is no particular need to get rid of the old concept.

In that sense, Scharp's theory agrees that in the vast majority of cases, ascending and descending truth overlap, which means that in practice we still have the T-schema. Even for Scharp's theory, which is a direct attack on the T-schema, the T-schema still appears as a fundamental feature of the common understanding of 'truth'.

4.4 Concluding remarks

Although I have not established a definite proof for the T-schema and cannot assert with certainty that the T-schema is a component of the everyday conception of truth, there is nevertheless evidence that supports taking the T-schema as a crucial minimal requirement for truth. Even while remaining as neutral as possible about what a full definition or theory of truth would look like, there is reason to think that the T-schema may very well be part of such a theory. At the very least, it is worth investigating what happens if the T-schema is kept and to explore non-classical options.

Chapter 4 – Three-valued logical systems

The liar paradox is caused by three features: a truth predicate satisfying the T-schema, a language whose sentences can refer to their own truth-values, and classical logic. The two preceding chapters have provided a justification for keeping the two first features; the second part of the thesis will now explore what happens in non-classical logic, starting with three-valued logical systems in this chapter. The first section of this chapter will show that two particular principles of classical logic cause the liar paradox, namely the law of excluded middle and the principle of explosion. The second section will present the type of logic, paracomplete logic, that emerges when the law of excluded middle is set aside, and the third section will introduce the logic without the principle of explosion, called paraconsistent logic. Finally, the last section will give a short account of how a truth predicate can be constructed in these logical systems, based on Kripke.

Section 1 – The liar paradox in classical logic

Classical logic is one of the three features that causes the liar paradox. But what exactly is it about classical logic that, when combined with the T-schema and natural language, gives rise to the liar paradox? Let us set up the paradox formally to analyse it in more detail.

Let λ be the liar sentence that says of itself that it is false. With the help of the truth predicate T introduced in the previous chapter, we can write $\lambda \coloneqq \neg\, T\langle\lambda\rangle$. We also have the T-schema applied to the liar sentence: $\lambda \leftrightarrow\ T\langle\lambda\rangle$.

In other words, we have:

Definition of λ	$\lambda \coloneqq \neg\, T\langle\lambda\rangle$
T-schema	$\lambda \leftrightarrow\ T\langle\lambda\rangle$

It is rather obvious that these two expressions, when taken together, will create an inconsistency. The liar sentence means "λ is not true", but it is also equivalent to "λ is true".

Let us set up the full argument in natural deduction.[67] First (1) we will assume that the liar sentence is true, from which we can deduce that it is not true. Thus λ is both true and not true. Conversely (2), if we assume that the liar sentence is not true, we can derive that it is not true.

[67] Adapted from Beall, Glanzberg and Ripley, *Formal theories of truth*, 21.

Then λ is again both not true and false. In either case (3), the liar sentence is both true and not true, and we have our contradiction.

(1) Assume:	1. $T\langle\lambda\rangle$		Premise
	2. λ	1	T-schema
	3. $\neg\, T\langle\lambda\rangle$	2	Definition of λ
	4. $T\langle\lambda\rangle \wedge \neg\, T\langle\lambda\rangle$	1,3	Conjunction principle
(2) Assume:	5. $\neg\, T\langle\lambda\rangle$		Premise
	6. λ	5	Definition of λ
	7. $T\langle\lambda\rangle$	6	T-schema
	8. $T\langle\lambda\rangle \wedge \neg\, T\langle\lambda\rangle$	5,7	Conjunction principle
(3) Then:	9. $T\langle\lambda\rangle \vee \neg\, T\langle\lambda\rangle \vdash T\langle\lambda\rangle \wedge \neg\, T\langle\lambda\rangle$	1,4,5,8	Disjunction principle
	10. $T\langle\lambda\rangle \wedge \neg\, T\langle\lambda\rangle$	9	Law of Excluded Middle, Closure
	11. B	10	Explosion principle, Closure

There is no restriction on what B is: it could be any sentence, such as "$2 + 2 = 5$", "The moon is made of camembert", or the craziest conspiracy theory of your choosing. Classical logic allows us to derive any false statement from the liar sentence. Where exactly do things go wrong?

The logical principles used in the derivation are as follows:

Conjunction principle	If $A \vdash B$ and $A \vdash C$, then $A \vdash B \wedge C$.
Disjunction principle	If $A \vdash C$ and $B \vdash C$, then $A \vee B \vdash C$.
Closure	If A and $A \vdash B$, then B.
Law of Excluded Middle	$\vdash A \vee \neg A$
Explosion principle	$A \wedge \neg A \vdash B$

The three first ones are fairly basic and uncontroversial. The conjunction and disjunction principles describe the behaviour of the operators ∧ (AND) and ∨ (OR). The closure principle governs the behaviour of the single turnstile ⊢ which indicates implication between two expressions. We will accept these three and instead focus on the two last principles.

The first of these is the law of excluded middle (LEM), according to which any sentence A is either true or false: $A \vee \neg A$. There is no "middle" or third option between truth and falsity; every sentence must either be true or be false. The second is the explosion principle (also called *ex falso quodlibe*t (EFQ) or *ex contradiction quodlibet* (ECQ)), which states that any arbitrary statement, including a false one, can be derived from a contradiction, that is, from a sentence being both true and false. Taken together, these principles mean that every statement must be either only true or only false.

The heart of the liar paradox thus lies with the law of excluded middle and the principle of explosion. Giving up on either of them keeps the paradox to emerge from the liar sentence. Although the three first principles can also be discussed, especially the closure principle, it would be quite onerous to abandon either of the three, and this would lead us towards logical systems that are much further removed from classical logic.

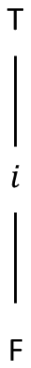
Abandoning these two principles opens up two types of solutions for the liar paradox: if the law of excluded middle is abandoned, we get paracomplete logics; without the principle of explosion, we get paraconsistent logics.

Section 2 – Paracomplete logic

Intro section somewhere strategy section: start with giving up LEM and seeing what kind of logic is necessary for that.

The first line of attack against classical logic will be to give up on the law of excluded middle $\vdash A \vee \neg A$. Statements will no longer have to be either true or false. This opens up for a family of logical systems called "paracomplete": a complete system will assign the truth-value 'true' or 'false' to every statement; without LEM, some statements can be neither one nor the other. As a result, the logic is no longer complete. However, this incompleteness is not an issue, hence the term "paracomplete". Since there now is a "gap" between the two classical truth-values true and false, this type of logic is also called "gappy".

Since every sentence no longer has to be truth or false, we introduce a third intermediary truth-value i.



The only requirement for this logic so far is that the law of excluded middle must not hold. Although one may intuitively think that any logic with a third truth-value will automatically violate the law of excluded middle, this is in fact not the case. To see this, a little technical machinery will be needed.[68]

Let $\mathcal{V}$ be the set of truth-values. In classical logic, this is the set $\{T, F\}$, but in paracomplete logic, $\mathcal{V} = \{T, i, F\}$. We also need an interpretation ν which maps any formula onto a truth-value in $\mathcal{V}$. For instance, if A is true and takes the truth-value T, we write $\nu(A) = T$.

Next, we have the set $\mathcal{C}$ of connectives $\{\neg, \wedge, \vee, \rightarrow\}$. For each connective c in $\mathcal{C}$, there is an associated truth function f_c that takes truth-values as input and output. In the case of negation $\neg$ for instance, $f_\neg(T) = F$. The meaning of the connectives is given by their truth functions.

These functions, the truth functions f_c and the interpretation ν can all be combined by function composition. The truth functions f_c take truth-values as input, and cannot therefore be directly applied to formulas or statements. It does not make sense to write $* f_c(p)$. First, we must use ν to find the truth-value of the formula, and then the truth functions f_c can be used.

For instance, suppose that we would like to know the truth-value of $p \vee \neg q$, given that $\nu(p) = T$ and $\nu(q) = F$ and we would like to know the truth-value of $p \vee \neg q$.

Then,
$$\begin{aligned} \nu(p \vee \neg q) &= f_\vee\big(\nu(p), \nu(\neg q)\big) \\ &= f_\vee\Big(\nu(p), f_\neg\big(\nu(q)\big)\Big) \\ &= f_\vee\big(T, f_\neg(F)\big). \end{aligned}$$

Using the truth-tables below,

[68] Terminology taken from Priest, Graham. *An Introduction to Non-Classical Logic: From If to Is*.

$$f_{\vee}(T, f_{\neg}(F)) \quad = f_{\vee}(T, T) = T.$$

These functions are easiest to represent by truth tables; for the sake of simplicity, we will write the connective itself and not its associated truth-function. Where both inputs are classical ($\{T, F\}$) the outputs are also the same as in classical logic.

Negation ¬

¬	
T	F
i	i
F	T

The rows for T and F are the same as for classical logic. If $f_{\neg}(i)$ was anything other than i, the law of double negation would no longer hold. Let $v(p) = i$, and suppose that $v(\neg p) = T$. Then $v(\neg\neg p) = F$ and we get $v(p) \neq v(\neg\neg p)$. Therefore $f_{\neg}(i)$ cannot be T. A similar argument rules out $f_{\neg}(i)$ having the value F. Thus $f_{\neg}(i) = i$.

Conjunction ∧

∧	T	i	F
T	T	i	F
i	i	i	F
F	F	F	F

As with classical logic, the conjunction is only true if both conjuncts are true. The conjunction is false whenever one of the conjuncts is false. In the three remaining cases, neither of these rules applies, and the conjunction gets the third intermediate value i.

Disjunction ∨

∨	T	i	F
T	T	T	T
i	T	i	i
F	T	i	F

The disjunction is only false when both disjuncts are false, and it is true wherever one of the disjuncts is true. In the three remaining cases, neither rule applies and the disjunction gets the intermediate value i. There is a symmetry between the truth tables for disjunction and conjunction and the De Morgan laws hold.

We can now look at a truth table for the law of excluded middle, using the truth tables for negation and disjunction

p	$\neg p$	$p \vee \neg p$
T	F	T
i	i	i
F	T	T

In order for the law of excluded middle not to hold, we need a counter-example. This can only be provided by the counter-model $v(p) = i$, since this is the only value of p for which $v(p \vee \neg p)$ is not T. In other words, we need this new third-value to be considered a counter-example disqualifying a formula from being considered true.

This is where the concept of "designated values" comes in. The set $\mathcal{D}$ of designated values is a subset of the set of truth-values $\mathcal{V}$: they are the truth-values that are preserved by valid entailments. In terms of classical logic, logic can be seen as preserving truth: the rules governing which entailments are valid provides rules to deduce true formulas from prior true formulas. Truth T is the only designated value in classical logic. In non-classical logic, however, designated values can be truth-values beyond T.

If we go back to our counter-example for the law of excluded middle, it means that for LEM not to be valid, there has to be a case where $v(p \vee \neg p)$ has a value that is not a designated one. Removing T from the list of designated values would require us to change our understanding both of the truth-value T itself and of what designated values are, and this would take us quite far from classical logic. The alternative is to consider the third truth-value i not to be designated. In that way, there is a case where $v(p \vee \neg p)$ does not take a designated value and the law of excluded middle no longer holds.

Abandoning the law of excluded middle while trying to remain as close as possible to classical logic thus takes us to a logical system with a third undesignated truth-value i.

We can check that the principle of explosion $p \wedge \neg p \vdash q$ still holds.

p	q	$p \wedge \neg p$	q
T	T	F	T
T	i	F	i
T	F	F	F
i	T	i	T
i	i	i	i
i	F	i	F
F	T	F	T
F	i	F	i
F	F	F	F

Since i is not a designated value, there is no case where $\nu(p \wedge \neg p)$ is a designated value and where q is not. Thus, there is no counter-model: the principle of explosion holds. The typical interpretation of the third-value i is as "neither true nor false", as it fills the "gap" between the two classical truth-values.

The paracomplete logic described so far is a common framework used by several logical systems. The most well-known is perhaps the Strong Kleene logic K_3. Its connectives are as described above, and implication $\rightarrow$ is defined through negation and disjunction as in classical logic: $p \rightarrow q$ is equivalent to $\neg p \vee q$. The truth-table for implication $\rightarrow$ is therefore as follows.

Implication $\rightarrow$	T	i	F
T	T	i	F
i	T	i	i
F	T	T	T

The following chapters will focus on the Strong Kleene logic K_3, and not other paracomplete logics, partly because it is the best-known paracomplete logic, and partly because its connectives are still defined in a way that is quite close to classical logic. However, other paracomplete logics exist and define connectives slightly differently. For instance, the law of

identity ($\vdash p \to p$) does not hold in K_3 (take $\nu(p) = i$). However, by defining implication a little differently, we can have a paracomplete logic in which the law of identity holds:

$\to$	T	i	F
T	T	i	F
i	T	T	i
F	T	T	T

The resulting logic was introduced by Łukasiewicz and is called $Ł_3$. There are also other paracomplete logics such as the Weak Kleene logic, which uses a similar framework as we have seen thus far, with the exception that in every truth-table, if the output is i then the output is i as well.

Section 3 – Paraconsistent logic

The second line of attack against classical logic is to give up on the principle of explosion $p \wedge \neg p \vdash q$.

For the principle of explosion not to hold, we need at least one case where the truth-value of $p \wedge \neg p$ is designated and the truth-value of q is undesignated. In other words, we need $\nu(p \wedge \neg p) \in \mathcal{D}$ and $\nu(q) \notin \mathcal{D}$.

We know that the truth-value T is a designated value in $\mathcal{D}$. However, requiring that $\nu(p \wedge \neg p) = T$, when p is an arbitrary formula, leads to some strange conclusions. If $p := 2 + 2 = 5$, it would mean that both $2 + 2 = 5$ and $2 + 2 \neq 5$ would be true. This would take us both quite far from classical logic and also quite far from our normal understanding of "true".

Instead, it is more natural to introduce a new truth-value that will be designated. Let us again call this third truth-value j. As with paracomplete logic, we have a set of truth-values $\mathcal{V}$ that is $\{T, j, F\}$ but the set of designated values $\mathcal{D}$ is now $\{T, j\}$. We can then build a counter-model with $\nu(p)$ such that $\nu(p \wedge \neg p) = j$ and $\nu(q) = F$.

The truth tables for negation $\neg$, conjunction $\wedge$ and disjunction $\vee$ are the same as in the paracomplete logic K_3 of last section, using the same type of justification.

Negation $\neg$

¬	
T	F
i	i
F	T

Conjunction ∧

∧	T	i	F
T	T	i	F
i	i	i	F
F	F	F	F

Disjunction ∨

∨	T	i	F
T	T	T	T
i	T	i	i
F	T	i	F

The table for the principle of explosion thus remains the same:

p	q	$p \wedge \neg p$	q
T	T	F	T
T	j	F	j
T	F	F	F
j	T	j	T
j	j	j	j
j	F	j	F
F	T	F	T
F	j	F	j
F	F	F	F

Now that the third truth-value is a designated value, we can find a counter-model: we need $\nu(p \wedge \neg p) = j$ and $\nu(q) = F$. Our counter-model is $\nu(p) = j$ and $\nu(q) = F$. Thus, the principle of explosion no longer holds.

We now have a logical system where $p \wedge \neg p$ is no longer automatically false but is instead assigned a designated value. In other words, it is no longer false that a sentence and its negation both be true, which goes against the consistency requirement of classical logic. This is why logical systems in which the principle of explosion does not hold are called "paraconsistent". The third truth-value is typically interpreted as "both true and false" in paraconsistent logic. Since the two classical truth-values seem to overlap and glut together, this type of logic is called "glutty".

We can check that the law of excluded middle holds when j is a designated value.

p	$\neg p$	$p \vee \neg p$
T	F	T
j	j	j
F	T	T

For every value of p we have $\nu(p \vee \neg p) \in \mathcal{D}$: the law of excluded middle holds.

The framework given so far is common to several paraconsistent logics, of which the most famous if Graham Priest's logic of paradox (LP).[69]

It uses the same truth-function for implication as Strong Kleene K_3, with implication defined through negation and disjunction as with classical logic

$\rightarrow$	T	j	F
T	T	j	F
j	T	j	j
F	T	T	T

As for paracomplete logics, there are other ways of defining the connectives, resulting in other types of paraconsistent logics. For instance, one issue with LP is that *modus ponens* no longer

[69] Priest, Graham. "The Logic of Paradox",

holds: $p, p \rightarrow q \nvdash q$. (The counter-model is $\nu(\mathrm{p}) = \mathrm{j}$ and $\nu(q) = F$.) This can be changed by adjusting the truth-table for implication:

$\rightarrow$	T	i	F
T	T	F	F
i	T	i	F
F	T	T	T

The resulting logic is the dialethic logic $RM3$ (which is the 3-valued extension of the logic R-mingle, itself an extension of the principal relevance logic R developed by Anderson and Belnap.)[70]

However, we will focus on the logic of paradox LP among paraconsistent logics in the remainder of the thesis for the same reasons that we will focus on the Strong Kleene logic K_3 among the paracomplete logics: their connectives are defined in a rather intuitive way that remains close to classical logic, and they are also the two best-known and most developed logical systems among the paracomplete and paraconsistent families.

Section 4 – Kripke construction in K_3

Now that we have a three-valued non-classical logic, we can construct an interpretation for a truth predicate for natural language that satisfies the T-schema and which deals with the liar paradox. Kripke showed in 1975[71] how to do this using the Strong Kleene logic K_3.

Kripke's starting point is a criticism of Tarski's truth theory, which he calls the "orthodox approach" and involves a hierarchy of languages and truth concepts (see chapter 2 section 3). For Tarski, a language cannot contain its own truth concept. A sentence cannot say something about the truth of another sentence in the same language. Tarski's solution is an object language without a truth predicate, and a metalanguage that contains the object language with the addition of a truth predicate used to express the truth-values of sentences in the object language. This construction can be repeated to establish an infinite layering of 'meta-metalanguages' with each its own truth predicate that can be applied to the language below. This way, a sentence cannot say of itself that it is false and the liar paradox is avoided.

[70] Anderson and Belnap, *Entailment: The Logic of Relevance and Neccessity*
[71] Kripke, "Outline of a theory of truth."

However, Kripke points out that in many instances it is perfectly unproblematic to have sentences that refer to the truth-value of another sentence, and that there's no need for metalanguage constructions. For instance, take the sentence 'Oslo is the capital of Norway'. This sentence is true; hence it is false to say that 'The sentence 'Oslo is the capital of Norway' is false'. This is not problematic and does not seem to require two different truth predicates. We can continue: 'The sentence ''Oslo is the capital of Norway' is false' is false' is true. Finally, we get 'It is true that the sentence 'The sentence ''Oslo is the capital of Norway' is false' is false''.

According to Tarski's theory, we need four languages and three different truth predicates to express this. Let's call the bottom layer, the object language, $\mathcal{L}_0$. This language does not have a truth predicate. Let $\mathcal{L}_1$ be the closest metalanguage which contains $\mathcal{L}_0$ and a truth predicate Tr_1, used to express the truth of sentences in $\mathcal{L}_0$. Similarly, we can then construct the 'meta-metalanguage $\mathcal{L}_2$ with the truth predicate Tr_2 that applies to sentences in $\mathcal{L}_1$ and so on.

Then 'Oslo is the capital of Norway' is a sentence in $\mathcal{L}_0$. 'The sentence 'Oslo is the capital of Norway' is false' can be reformulated as $\neg Tr_1('Oslo\ is\ the\ capital\ of\ Norway')$ and is a sentence in $\mathcal{L}_1$. In order to express the final sentence, 'It is true that the sentence 'The sentence ''Oslo is the capital of Norway' is false' is false'', we need the meta-meta-metalanguage $\mathcal{L}_3$ and its associated truth predicate: $Tr_3(\neg Tr_2(\neg Tr_1('Oslo\ is\ the\ capital\ of\ Norway')))$.

Kripke points out that his kind of strategy is perhaps a little excessive and unnecessary when it comes to sentences that do not include semantic paradoxes. When we know the truth-value of the basic sentence, which in this case is 'Oslo is the capital of Norway', we know the truth-value of all the sentences using this as a building block, no matter how many "it is true that" or "it is false that" we attach to it. There is no need to have a different truth predicate for each step.

Kripke calls these types of sentences "grounded": their truth-value can be decided by looking at other sentences by a process that "terminates in sentences not mentioning the concept of truth"[72]. Sentences that are not grounded are called "ungrounded". The truth-teller sentence ('This sentence is true') is a typical example of an ungrounded sentence: although it does not create a paradox, its truth-value is impossible to determine, as its truth only depends on itself.

The main strategy used by Kripke is to construct a truth-predicate that will assign the value 'true' to true grounded sentences, 'false' to grounded false sentences, and the third truth-value

[72] Kripke, "Outline of a theory of truth.," 694

i to the ungrounded sentences. There is thus no need for a Tarskian infinite hierarchy of languages and sentences that cause semantic paradoxes are taken care of.

In the following we will give a sketch of Kripke's methodology. Let us start with the universe $\mathcal{U}$ of all sentences.

We will introduce a truth predicate Tr by constructing its extension E and antiextension A. The extension E will be the set of true sentences and the antiextension A the set of false sentences. The truth predicate will be well-defined once the extension and antiextension are established.

In a classical framework, all sentences are either only true or only false, so establishing the extension of the truth predicate would be sufficient. All elements not in the extension are automatically part of the antiextension. In the three-valued Strong Kleene logic used by Kripke, however, some sentences are neither true nor false but instead have the third truth-value i. The extension E and antiextension A cannot overlap but there may be sentences that are in neither of them: there will be a gap between E and A. This is why it is necessary to construct both the extension and the antiextension for the truth predicate to be well-defined.

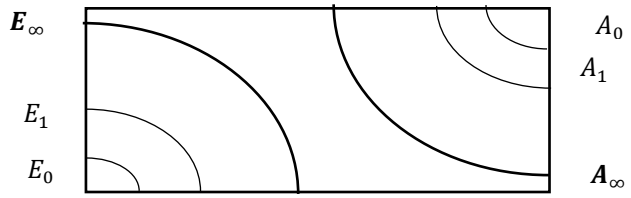
We will start with both the extension E and the antiextension A being empty: let us call them E_0 and A_0 at that stage.

For the next stage, we can fill the extension of E_0 with all the basic sentences that we know are true and which do not contain any reference to the truth or falsity of another sentence. This truth predicate satisfies the T-schema: sentences such as 'Dogs are mammals' are taken to be true if and only if dogs are indeed mammals. This new extension is called E_1. Similarly, we construct A_1 by filling A_0 with all the sentences we know are false that do not refer to the truth-value of other sentences. At this stage, sentences such as 'Susan Stebbing was a logician' end up in E_1 and 'Oslo is the capital of France' in A_1.

For the next stage, we determine the truth-value of more sentences in order to construct E_2 and A_2, by building on E_1 and A_1. First, all sentences already in E_1 or A_1 are automatically part of E_2 or A_2 respectively. Next, any sentence that asserts the truth or falsity of a sentence already established to be in E_1 or A_1 can easily be determined to be either true or false: if it asserts the truth of a sentence in E_1 or the falsity of a sentence in A_1, it is true and can be put in E_2; if it asserts the falsity of a sentence in E_1 or the truth of a sentence in A_1, it is false and can be put in A_2. The sentences 'It is true that Susan Stebbing was a logician' and 'It is false that Oslo is the capital of France' thus both end up in E_2.

In this manner, we can continue to build E_3, A_3, E_4, A_4, … and continue into the transfinite. At limit stages, we take the union of all the preceding E_n for n in ℕ, and call this E_ω, where ω is the smallest infinite ordinal number. We can then continue, repeating the same procedure, with $E_{\omega+1}, E_{\omega+2}, \ldots$ and so on. The same process happens for the antiextensions. In this way, all grounded sentences eventually end up in E_α or A_α for some ordinal α.

If a sentence has been shown to be true or false at a particular step, that is, it is part of E_α or A_α for some ordinal α, it will not change truth-value at a later step. In other words, it will remain in E_β or $A_{\backslash beta}$ for all ordinals β such that $\beta \geq \alpha$. This establishes monotonicity of the sequences $E_0, E_1, \ldots, E_\omega, \ldots$ and $A_0, A_1, \ldots, A_\omega, \ldots$. Kripke thus proves that eventually both the extension E and the antiextension A will reach a fixed point E_∞ and A_∞ and that a gap between them will remain, containing the ungrounded sentences, including the liar sentence, which get assigned the third truth-value i.



This way Kripke builds a truth predicate defined by its extension and antiextension. All grounded sentences have been assigned a truth-value 'true' or false', and the ungrounded sentences are assigned the third truth-value i, which can here be interpreted as 'unknown' or 'undefined' and as being neither true nor false.

Kripke did his construction using Strong Kleene K_3, but a similar proof can be made with other three-valued logics, including the paraconsistent logic of paradox LP. For other paracomplete logics a similar strategy can be used, but for a paraconsistent logic such as LP the strategy is "reversed". The starting point is to take E_0 and A_0 to be the entire universe $\mathcal{U}$ of all sentences (instead of taking them to be empty). Then, at the first stage, the truth-value of basic sentences is established, using the T-schema, so that true sentences such as 'Anne Conway was a philosopher' are removed from the antiextension, but remain in E_1, and false sentences are removed from the extension but remain in A_1. Thus E_1 and A_1 are both smaller than E_0 and A_0. At the next stage, E_2 and A_2 are established by removing from E_1 and A_1 sentences that express the truth-value of the preceding basic sentences. For instance, the sentence 'It is true that Anne Conway was a philosopher' remains in E_2 but not in A_2. This construction continues until a fixed point, but this time E_∞ and A_∞ will overlap (and not have a gap between them). The sentences in E_∞ are assigned the truth-value 'true', the sentences in A_∞ are assigned 'false', and the sentences in the overlapping area, liar sentence included, are assigned the third truth-value j, which will typically be interpreted as 'both true and false'.

In this way, both K_3 and LP can be used to construct a truth predicate satisfying the T-schema in natural language.

Chapter 5 – Towards a unifying framework

In order to have a truth predicate satisfying the T-schema in natural language, it is necessary to give up on either the law of excluded middle or the principle of explosion, which leads to two families of solutions, paracomplete logic with K_3 and paraconsistent logic with LP.

At first glance, the Strong Kleene logic K_3 and the logic of paradox LP are quite similar. Both are adaptations of classical logic; both have a third truth-value in addition to the classical 'true' and 'false'. Furthermore, both can be used in Kripke-style construction of a truth predicate and their truth-tables are identical. Is one of them a better framework than the other for modelling truth? If not, is the solution to be pluralistic about truth?

This chapter will clarify the differences between the two third truth-values i (in K_3) and j (in LP) (sections 1 and 2), and will argue that K_3 and LP each have their own preferred area of application (section 3). However, instead of advocating for pluralism, a unifying framework can be found through the four-valued logic FDE (section 4).

Section 1 – The many interpretations of a third truth-value

Many-valued logics have a fairly long history, and quite a few three-valued logical systems have been developed. Typically, the newly introduced third truth-value has been understood as meaning something along the lines of "intermediate", "neutral", or "indeterminate".[73] However, there has been a lot of discussion about how exactly to understand and to interpret this third truth-value. The purpose of this section is not to be historically exhaustive in any way, but to give an idea of the wide array of meanings the third truth-value can take and to point out some of the disagreements surrounding these issues.

The first to introduce a new truth-value beyond the two classical ones was Łukasiewicz in 1920, motivated by possible propositions that express something that may or may not happen in the future. He considers the proposition "I shall be in Warsaw at noon on 21 December of next year"[74], which expresses a possibility, but not a necessity. It can neither be true nor false: if it were true, Łukasiewicz's future presence in Warsaw would be necessary; if it were false, it would be impossible. Neither case can be right, so the proposition can be neither true nor false. Therefore, he introduces a three-valued logic now called $Ł_3$, with a third truth-value designated

[73] Rescher, *Many-Valued Logic*, 22.
[74] Łukasiewicz, 'Philosophical remarks on many-valued systems of propositional logic,' 53

by ½ ('true' is typically seen as '1' and 'false' as '0') to represent "the possible"[75], and to be assigned to propositions that are neither true nor false.

This interpretation was disputed by Moh Shaw-Kwei, who meant that the third truth-value should not be used for possible propositions about the future. His argument rests on the simple issue that a sentence is not identical to its negation: to use Łukasiewicz's example, "I shall be in Warsaw at noon on 21 December of next year" is not the same as, "I shall not be in Warsaw at noon on 21 December of next year", yet the two propositions have the same truth-value in $Ł_3$. Instead, the only propositions that should be assigned the third truth-value ½ are propositions that are equivalent to their own negation, that is, paradoxical propositions.[76]

In 1938, Kleene introduced a third truth-value in a paper on ordinal numbers[77], in the context of partial recursive functions. For some sentences containing such functions, there is no algorithm that can be applied in order to determine whether the sentence is true or false. Kleene therefore introduced the value 'u' for 'undefined'. Kleene developed this further in his 1952 book *Introduction to Metamathematics*[78], where he set up a fully-fledged three-valued logic (the Strong Kleene logic K_3) with the three truth-values 't', 'f' and 'u' for 'true', 'false' and 'undefined' respectively. However, he made it clear that it must be possible to interpret the third truth-value differently than just 'undefined', for instance as "unknown (or value immaterial.)"[79] This new interpretation makes it possible for any proposition to have a third truth-value, beyond algorithmically undetermined sentences in the context of partial recursive functions. The value 'unknown' is meant to apply for any statement where it is not known (or deliberately disregarded) whether it is true or false: he writes that "u means only the absence of information"[80] whether the statement is true or false, and can therefore be understood in different ways.

Goddard and Routley also note that Kleene seems to infer that the third-value 'u' could also be interpreted as 'meaningless', [81] which they take to be equivalent to "nonsignificant"[82]. However, they argue that such an interpretation is not legitimate, since a nonsignificant

[75] Ibid.
[76] Shaw-Kwei. "Logical Paradoxes for Many-Valued Systems."
[77] Kleene, "On Notation for Ordinal Numbers," 153.
[78] Kleene, *Introduction to Metamathematics*
[79] Ibid., 335
[80] Ibid.
[81] Goddard and Routley, *The Logic of Significance and Context*, 266; referring to Kleene, *Introduction to Metamathematics*, example 1 p.335
[82] Ibid.

sentence does not have a truth-value, whereas Kleene's third value is intended to represent a lack of knowledge about the sentence's truth-value. In other words, Kleene's third value merely represents an epistemic state: the sentence is either true or false, but it is not known which one. In contrast, a 'meaningless' or 'nonsignificant' sentence lacks a truth-value altogether ontologically.[83]

The same year as Kleene's first paper, in 1938, Bochvar also proposed a three-valued logic, called B_3.[84] Rescher takes the third truth-value to mean "undecidable"[85], which seems to be very close to Kleene's algorithmic undecidability. However, in B_3, the third value (called S) means "having some element of undecidability about it,"[86] and if some element of the sentence is undecidable, the entire sentence becomes undecidable. (This logic is the same as Weak Kleene, mentioned in chapter 4.) For instance, if $v(p) =_{B_3} T$ and $v(q) =_{B_3} S$, then $v(p \vee q) =_{B_3} S$. In other words, the undecidable element transmits its undecidedness to whichever sentence it is a part of, even when this means that one true disjunct is no longer sufficient to make the disjunction true. In contrast, in Kleene's K_3, if $v(p) =_{K_3} T$ and $v(q) =_{K_3} u$, then $v(p \vee q) =_{K_3} T$. This means that whether one interprets the third truth-value as "undefined" or "undecidable" is not just a matter of semantics, but has important consequences for the construction of the connectives. However, I noted that Bochvar's original article in Russian uses the term "бессмыслица"[87] (bessmyslitsa), which means "nonsense", and that is indeed the translation chosen in Bochvar's English article from 1981. This third value has also been interpreted as "paradoxical" and "meaningless"[88].

Another three-valued logic which interpreted the third value as meaningless or 'nonsense' is Logic of Nonsense introduced by Halldén[89] in 1949. Halldén was the first to introduce a third truth-value that was designated.[90] Another three-valued logic with a designated value was introduced by Ulrich Blau[91] in 1977, with the purpose of having a logic that best formalised the

[83] Szmuc and Omori. "A Note on Goddard and Routley's Significance Logic," 434.
[84] Bochvar, "On a three-valued logical calculus and its application to the analysis of contradictions"
[85] Rescher, *Many-Valued Logic*, 29.
[86] Ibid.
[87] Bochvar, "On a three-valued logical calculus and its application to the analysis of contradictions," 289.
[88] Rescher, *Many-Valued Logic*, 29.
[89] Hallden, "The logic of nonsense,".
[90] Omori, "Hallden's Logic of Nonsense and Its Expansions in View of Logics of Formal Inconsistency," 3
[91] Blau, *Die dreiwertige Logik der Sprache: ihre Syntax, Semantik und Anwendung in der Sprachanalyse*.

features of natural language. Blau's third value is called "unbestimmt", that is, "undecided" and is meant to be applied to sentences with vague concepts and denotation failures.[92]

Finally, Graham Priest introduced the logic of paradox *LP* in 1979[93] with the purpose of making paradoxical statements acceptable. The third truth-value in *LP* is designated and is meant to be assigned to paradoxical statements such as semantic paradoxes, that according to Priest are "true contradictions". This is a similar interpretation of the third truth-value to Shaw-Kwei's proposal, although Shaw-Kwei simply interpreted the undesignated third value in $Ł_3$ as 'paradoxical'. By making the third truth-value designated, it is preserved by logical entailment and it plays therefore in some sense a "truth-like" role. A typical interpretation of Priest's third truth-value is "both true and false".

Thus, in the rich history of three-valued logics, there has been a lot of discussion about the meaning of the third truth-value: should it have a modal character to denote possibility or future contingent propositions? Should it be understood epistemically, and be assigned to sentences whose truth-values are unknown, or ontologically, and be used only for sentences that are really neither classically true nor classically false? Should it mean undefined, undecidable, unprovable, unknown, meaningless, nonsensical, nonsignificant, paradoxical?

Section 2 – Designated and undesignated values

In order to avoid these debates, I propose to focus less on the meaning and significance of the third truth-value, and instead consider its role more structurally. The more relevant question about the third truth-value should be whether or not it is designated, and not what its interpretation should be. Let us recall that a designated truth-value is a truth-value that is preserved by logical entailment; the purpose of logical entailment is to preserve the designated truth-values. In classical logic, 'true' is the only designated value, so it is natural to think of the role of entailment as being truth-preserving. It is perhaps therefore more intuitive, when introducing a third truth-value, to think of it as undesignated, so that entailment remains solely truth-preserving; most three-valued logics do indeed have an undesignated third truth-value in addition to the two classical ones. However, it is also possible to introduce a third truth-value that is designated, as Halldén, Blau and Priest have done. Designated values can then no longer be identified with "true"; instead, "true" is merely one of several designated values.

[92] Blau p.21
[93] Priest, "The Logic of Paradox."

By giving more importance to whether the truth-value is designated or undesignated, the discussion about the meaning of the third truth-value becomes less important. Defining the third value more structurally, instead of focussing on its intended meaning and application, lets the third truth-value remain open to more interpretations. The Strong Kleene logic K_3, for instance, was created with a more epistemic understanding in mind, but there is no reason why an ontological interpretation cannot be given. The same goes for the logic of paradox LP, which was intended for true contradictions, that is, for propositions that ontologically are both true and false. However, LP can also be interpreted epistemically and used for propositions where there is information about them being true but also information that they are false.

Obviously, there are some limitations depending on how the connectives are defined and each three-valued logic cannot be given every single possible interpretation. Weak Kleene logic, for instance, does not make much sense if the third truth-value is interpreted as 'lack of knowledge'. If it is known that p is true and that the truth-value of q is unknown, it is more intuitive to define disjunction so that 'p or q' is known to be true, since one of the disjuncts is known to be true. In Weak Kleene logic, however, if $v(p) = T$ and $v(q) = u$, then $v(p \vee q) = u$, so an interpretation along the lines of 'meaninglessness' makes more sense than 'lack of knowledge'.

In the rest of the thesis, I will continue with the choice made in chapter 4 to focus on Strong Kleene K_3 and on the logic of paradox LP.

Section 3 – K_3 and LP

The most common interpretation of Strong Kleene K_3 is that it creates a gap between 'truth' and 'falsity', hence its denotation as 'gappy logic'. The third truth-value is therefore seen as being between 'true' and 'false' and is therefore typically interpreted as 'neither true nor false'. The opposite happens with the logic of paradox LP, where the two classical truth-values are seen as overlapping and create a glut, hence why it is called 'glutty logic'. The third truth-value denotes the 'area' where 'true' and 'false' overlap, and is therefore often interpreted as 'both true and false'.

However, this understanding of the new truth-values as meaning 'neither true nor false' and 'both true and false' can be "seriously misleading"[94], and this interpretation must not become a constitutive part of K_3 and LP. To see why this is, consider a proposition p such that $v(p) = i$. If p is neither true not false, then we can write $\neg(p \vee \neg p)$. However, according to the De

[94]Beall, Glanzberg and Ripley. *Formal theories of truth*, 41.

Morgan laws, this is equivalent to $(\neg p \wedge \neg\neg p)$, i.e., $(\neg p \wedge p)$, which says that p is both true and false. In other words, there is a formal equivalence between 'neither true nor false' and 'both true and false'. Despite this, and despite K_3 and LP sharing truth-tables, they are very much two different logical systems, and the importance of the third truth-value being designated or not cannot be overstated.

Both K_3 and LP can be interpreted in many different ways, which is why I prefer to denote the third-value as i in K_3 and as j in LP, as these are more neutral than 'u' ('undefined', 'unknown'), or than 'b' (both) and 'n' (neither).

There are areas of application where K_3 is more appropriate, and others where LP is preferred, even without relying exclusively on the intuitive 'neither/both' interpretation.

Two main motivations for the development of paracomplete logic are future contingents and epistemic lack of knowledge. Future contingents are statements about the future that may or may not happen: they must neither be inevitable nor impossible. This is Łukasiewicz's possible future propositions that motivated his development of $Ł_3$. A well-known example of a future contingent is found in Aristotle: "Tomorrow there will be a sea-battle"[95]. There might be a sea-battle or there might not be a sea-battle tomorrow and there is no way of knowing which. It can be seen both as an epistemic lack of information and as metaphysically undetermined. The two classical truth-values cannot be assigned to the statement, so a third truth-value is preferred. There is a case to be made that the statement being intuitively neither true nor false, this third truth-value assigned to the statement should be interpreted as 'neither true nor false', and should therefore be undesignated, following the rules of paracomplete logic. However, it is not necessary to embrace the intuitive interpretation to achieve this. Let S be the statement "Tomorrow there will be a sea-battle". This statement S cannot be asserted, but neither can its negation $\neg S$. We do not have $S \vee \neg S$; in other words, the law of excluded middle is violated. This is why paracomplete logic is preferred for future contingent statements, and as seen in chapter 4, without the law of excluded middle, the third truth-value must be undesignated.

The same happens for statements for which there is no information, such as Kleene's algorithmically undecidable statements that motivated the development of K_3. Such statements can neither be asserted nor denied, which violates the law of excluded middle and justifies the assignment of an undesignated third truth-value.

[95] Aristotle, *On interpretation*

Graham Priest's motivation for developing the logic of paradox LP was the notion of "true contradictions" and the presence of inconsistent laws. Consider a (hypothetical)[96] country where mothers who give birth are entitled to maternity leave. This is implicitly understood as meaning that maternity leave is 'maternal' and therefore for mothers, and that giving birth creates a specific benefit in the workplace. Now let us take the case of a transgender man who gives birth. Is he entitled to maternity leave? On the one hand, he is not a woman, and is therefore not entitled to maternity leave. On the other hand, he is giving birth, and is therefore entitled to maternity leave. Let T be the statement "A pregnant man is entitled to maternity leave". Both T and $\neg T$ can be asserted: we have $T \wedge \neg T$. In other words, we must avoid the principle of explosion, which leads us to prefer paraconsistent logic. From chapter 4 we know that three-valued logic without the principle of explosion has a designated truth-value, which is why inconsistent statements should be assigned a designated truth-value.

This is still consistent with the intuitive reading of the undesignated value i as 'neither true nor false' and of the designated value j as 'both true and false': future contingent statements seem to be neither true nor false while inconsistent statements seem both true and false. Even though there is a formal equivalence between 'neither true nor false' and 'both true and false', these interpretations are very intuitive and should not be considered to be wrong. It is simply necessary to be careful and to be aware that this interpretation can be misleading in certain contexts. Above all, it is important that the third truth-values i and j in K_3 and LP be primarily defined as undesignated and designated respectively, and not as 'neither true nor false' and 'both true and false'.

The important thing here is that although K_3 and LP are both three-valued with identical truth-tables, their respective extra truth-values are not the same, which leads the two logical systems to be very different in character. Depending on the area of application, either K_3 or LP might be preferred. In some cases, a designated third truth-value is needed, and in others an undesignated one is better.

Section 4 – Pluralism or a unifying framework?

Is either one of these logics better for modelling truth? There is an intuitive sense that the truth-teller sentence "This sentence is true" is best understood in K_3 whereas the liar sentence "This sentence is false" is best understood in LP. The justification for this is that the truth-teller is

[96] Following Graham Priest's example of an aboriginal property-holder in *An Introduction to Non-Classical Logic: From If to Is*, 128.

vacuously circular: nothing can be said about it. If it is true then it is true; if it is false then it is false. There is no justification for arguing one way or the other. It can therefore seem natural to see it as being neither true nor false. The liar sentence on the other hand is true if it is false and false if it is true: it seems to be both true and false. It can therefore be tempting to assign K_3's i to the truth-teller and LP's j to the liar sentence.

However, this can be disputed. Let λ be the liar sentence as previously, and let τ be the truth-teller sentence. For the liar sentence, if λ is true then λ is false, and if λ is false then λ is true. Similarly, for the truth-teller, if τ is true then τ is true, and if τ is false then τ is false. On the one hand there seems to be no basis on which to assert the truth or falsity of either λ and τ, they can both be considered as being 'neither true nor false'. On the other hand, there is just as much reason to argue that they are true as there is to argue that they are false, and as such λ and τ can both be considered as being 'both true and false'.

We know from the previous chapter that K_3 as well as LP both provide a solution to the liar paradox, the first by avoiding the law of excluded middle and the second by abandoning the principle of explosion. The truth-teller sentence does not give rise to a paradox like the liar sentence does, but as we have just seen, K_3 and LP are both capable of handling it.

Therefore, the two logical systems K_3 and LP do equally well in managing the liar paradox, but there are contexts such as future contingents where K_3 is preferred and others such as inconsistencies where LP makes more sense.

So far, we have not seen any criteria that allows us to prefer a paracomplete solution over a paraconsistent one. Should we therefore argue for some kind of neutral pluralism where K_3 and LP are both considered equally valid for modelling truth in natural language?

In a pluralistic perspective, K_3 and LP are considered as equally valid. For pluralists, there is no 'One True logic,' instead there can be several correct logics. One solution would therefore be to argue that K_3 and LP are both correct, sometimes, for example when dealing with future contingent statements, K_3 is the correct logic to use, while at other times, for instance to deal with inconsistencies, LP is the right logic to use.

There is something slightly unsatisfying about this solution: it can seem like an easy way out that avoids the responsibility of taking a position. Furthermore, we have already seen that there are contexts in which K_3 works best and others where LP is preferred, which means that this would not be the kind of pluralism for which all the theories are equally valid.

Perhaps more importantly, in classical logic, the law of excluded middle and the principle of explosion are two fundamental laws of logic that are considered crucial. In Strong Kleene K_3, the law of excluded middle is abandoned, but the principle of explosion remains just as strong and important as in classical logic. Similarly, in the logic of paradox LP, the principle of explosion is discarded while the law of excluded middle retains its critical and essential position.

If I were to embrace pluralism, K_3 and LP would be equally as valid, which would mean that their logical principles would be equally as fundamental. In other words, the principle of explosion in K_3 would be considered just as valid and essential as the principle of excluded middle in LP. However, conversely, K_3 and LP make equally valid choices when it comes to discarding the principle of excluded middle and the principle of explosion respectively. There is therefore a tension between these principles simultaneously being fundamental laws of logic, while also being easily discarded in the right context.

A solution out of this is a logical system that provides a unifying framework for both K_3 and LP. This is the four-valued logic called First Degree Entailment, or FDE, which will be the focus of the next chapter. In FDE, both the law of excluded middle and the principle of explosion are abandoned. The system has both additional values from K_3 and LP, which gives it four truth-values in total: t, f, i and j. Although FDE is necessarily weaker than both K_3 and LP, it can be considered a unifying framework for a 'pluralism in disguise.' The idea is that in the contexts where, for instance, it is applicable to use the law of excluded middle, this law can be added as a useful rule for a specific context. Similarly, a rule of explosion can be added in contexts where this may be useful. In this way, FDE can easily be strengthened toK_3, LP and classical logic. By demoting the law of excluded middle and the principle of explosion from fundamental laws of logic to mere rules, we avoid the tension plaguing pluralism between K_3 and LP. Furthermore, FDE allows us to move in contexts where both the law of excluded middle and the principle of explosion ought to be abandoned. However, the difficulty the pluralist faces in determining when to use K_3 and when to use LP remains to some extent, in FDE one must still determine when it is appropriate to add the additional strengthening rules. Nevertheless, this issue is less critical in FDE than for the pluralist.

Chapter 6 – FDE

In this chapter I present the logic FDE. The first section introduces the main elements of the logic while section 2 provides three motivations fpr FDE that will hopefully provide some intuitive understanding of the logic. In section 3 I present the four different semantics for FDE. The logic FDE is famous for not having an agreed-upon implication, so in section 4 I present a few options and argue in favour of an implication from relevant logic. Finally, in section 5 I define the T-schema in FDE; indeed, the entire point of this thesis was to have a truth concept with a T-schema. In the last section I show how FDE provides a solution to the liar paradox.

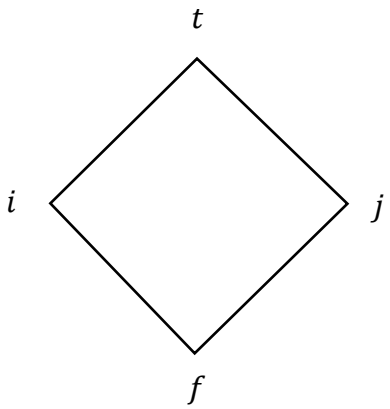
Section 1 – Introduction

The logic of First-Degree Entailment (FDE) is a four-valued logical system that is both paracomplete and paraconsistent (the term 'paranormal' is used for logics that are both paracomplete and paraconsistent). Neither the law of excluded middle nor the principle of explosion is valid in FDE, and it can be seen as a 'combination' of K_3 and LP.

The four truth-values are the classical two values 'true' ('t') and 'false' ('f'), with two additional intermediary values, one undesignated and one designated. These two intermediary values are those from K_3 and LP, and we will keep the notation 'i' for the undesignated value and 'j' for the designated one. The truth-values in FDE are often referred to as 'true', 'false', 'both' and 'neither,' and are symbolised with the sets $\{t, f, b, n\}$ or $\{1,0, b, n)$; but in order to avoid commitment to the 'both/neither' interpretation, I will prefer the more neutral symbols $\{t, f, i, j\}$.

A simple way to visualise FDE is with a lattice. The truth-values of K_3 and LP can easily be represented linearly, with i and j as intermediate values between 'true' and 'false'.

The two intermediate values of K_3 and LP are very different, as we saw in the previous chapter, and incommensurate. The four truth-values of FDE cannot be represented linearly. If one were to attempt that, how should one order i and j? If one follows the intuitive interpretation 'both true and false' and 'neither true nor false', it is clear that it makes no sense to wonder whether 'both true and false' is closer to 'true' or to 'false', and the same goes for 'neither true nor false'. One might think that j, since it is a designated value, is closer to t than to f, and that the undesignated value i is closer to f than to t, but this is misleading. The intermediate values cannot be compared, which is why the best way to represent FDE's four truth-values is as follows:



This lattice is a Hasse diagram, which is used to represent partially ordered sets. How the set truth-values $\{t, f, i, j\}$ can be analysed as being partially ordered will be discussed further in the context of the algebraic semantics of FDE (section 3.3).

Section 2 – Making sense of FDE

The logic FDE was developed by the logicians Nuel Belnap, Alan Ross Anderson and J. Michael Dunn in the 1950s and 1960s, and was motivated by concerns in relevant logic. A second motivation emerged in the 1970s to help computers process databases. Furthermore, Graham Priest has also argued that FDE is well-suited to formalise a certain Buddhist logic.

It can be difficult at first to make some intuitive sense out of the lattice diagram, but the epistemic interpretation provided by the database motivation is probably the most accessible, which is why I will start with start with that one instead of doing a chronological presentation.

2.1 Databases and computer science

Classical logic is not well-suited to handle databases. Imagine a large database containing hourly temperatures measured at Blindern in Oslo from 1950 until now. The database would be

the work of many different scientists, contain over 600,000 entries, and it is easy to imagine how some small errors may occur: contradictory information might be entered, or there could be missing information.

Suppose there is a missing temperature for 11 a.m. on the 17th of March 1978. What happens to the sentence S: "The temperature at Blindern at 11 a.m. on the 17th of March 1978 was 3.2°C"? Is S true or false? In the 'real world' so to speak, this sentence does of course take one of the two classical truth-values, but not within the database, where there is simply no information on the matter. Neither S nor $\neg S$ can be affirmed, and we find ourselves in a situation where the seemingly innocuous database is violating one of the most fundamental laws of classical logic, namely the law of excluded middle.

Conversely, suppose that two scientists both made an entry for the same time: one entered the value 17.2°C for 3 p.m. on the 21st of September 1997, another entered the value 17.3°C. Now consider the sentence S': "The temperature at Blindern at 3 p.m. on the 21st of September 1997 was 17.3°C". According to one line of the database it is true, but according to another it is false. What is the truth-value of S'? Both S' and $\neg S'$ can be affirmed, and with the help of the principle of explosion the database can now be used to prove any crazy conclusion, for instance that penguins are going extinct because they fall off the edge of the (flat) Earth.

This is why Nuel Belnap[97] pointed out that a four-valued logic without the law of excluded middle and the principle of explosion, such as FDE, is much better suited for computer science.

There are four states the computer can be in:[98]

T: the computer has been told that the statement is true

F: the computer has been told that the statement is false

N: the computer has not been told anything about the statement

B: the computer has been told that the statement is true and also that the statement is false

These four states T, F, N, B correspond to the interpretation 'true, false, neither, both' and thus to the truth-values t, f, i, j of FDE. In our hypothetical temperature database, the sentence S can now be assigned the truth-value i and the sentence S' can be assigned j, and the database can

[97] Belnap, "A useful four-valued logic"

[98] Based on Belnap, "A useful four-valued logic", 11 and Pietz and Rivieccio. "Nothing but the Truth," 127

no longer be used to derive absurd conclusions. The logic FDE is therefore much better suited than classical logic for computers to process information in databases.

2.2 Relevant logic

The logic of FDE was not in fact developed to deal with paradoxes and databases, but emerged in the context of relevant logic. Belnap first introduced FDE as E_{fde}, a fragment of the logic E in relevant logic[99] which imposes a further constraint on implication: the antecedent and the consequent must have something to do with one another, their subject matter must be linked somehow. We will see how FDE emerges in such a context.

Implication in classical knowledge can appear suspect. This is because material implication $A \supset B$ is defined as $\neg$A $\vee$ B. The truth-value of the implication depends exclusively on the truth-values of the propositional formulae: a false antecedent or a true consequent will make the implication true, regardless of the rest of the formula. This can have some odd results:

(p) Global warming is a hoax spread by polar bears

(q) Hypatia of Alexandria loved whisky

These two statements are not connected in any way, but, because (p) is false, '(p) implies (q)' is true. This seems odd: polar bears spreading misinformation is completely unrelated to Hypatia's drinking preferences. One might instead wish for the consequence relation to capture some kind of relation between the antecedent and the consequent, so that p implies q only when p has something to do with q. As David C. Makinson puts it: "no implication should hold in virtue of its antecedent alone, nor its consequent alone, but in virtue of a link between them."[100] Some logicians have therefore defended a consequence relation where the antecedent must be relevant to the consequent; the logical systems subsequently developed are called relevant logics (or relevance logics in the US).

Two principles in particular involve premises that are entirely irrelevant to the conclusion:

Law of Excluded Middle $B \vdash A \vee \neg A$

Explosion principle $A \wedge \neg A \vdash B$

[99] Omori and Wansing. "40 years of FDE: An Introductory Overview," 1021.
[100] Makinson, *Topics in Modern Logic*, 26.

As we saw in chapter 4, these two principles can be avoided by introducing two new truth-values, one designated and one undesignated.

Relevant logicians want to go further, however, and ensure that the consequence relation only holds when the antecedent and the consequent are on "the same topic". To be "on the same topic" is quite imprecise, but it is captured by the 'variable sharing principle', which states that the antecedent and the consequent must have at least one propositional variable in common for there to be a consequence relation between them.[101]

The logic FDE does in fact satisfy the variable sharing criterion. The proof can be found in Makinson's book[102] and in the subsection 4.2 of this chapter, and uses the algebraic semantics for FDE which I will introduce in the subsection 3.3.

2.3 Buddhist logic

Although FDE was developed with relevant logic in mind and later for computer science, Graham Priest[103] has argued that this logical system fits quite well with a type of Buddhist logic, which shows that a four-valued system is not as unintuitive as may first appear.

One of the earliest complete collections of Buddhist texts is the Pāli Canon, where a logical system of four alternatives can be found, known as the 'catuṣkoṭi' (or as 'tetralemma' in the West). Whereas in traditional Western logic (following Aristotle), a proposition is either true or not, there are four possibilities in the catuṣkoṭi. An example from the *Dīgha Nikāya*, as presented by Jayatilleke[104], goes as follows:

(1) A person is wholly happy
(2) A person is wholly unhappy
(3) A person is both happy and unhappy
(4) A person is neither happy nor unhappy

This may in fact appear closer to our own emotional experiences than the bivalent system of classical logic which has only two possibilities: either a person is happy or they are unhappy. Most of us have probably experienced feeling simultaneously happy and unhappy, or being in

[101] Mares, "Relevance Logic."
[102] Makinson, *Topics in Modern Logic*, see exercise 48, p.33 and p.94.
[103] Priest, "The logic of the catuṣkoṭi."
[104] Jayatilleke, K. N. "The Logic of Four Alternatives,", 70

an emotional state that is captured neither by 'happy' nor by 'unhappy', and one can argue that this four-fold system captures a lived experience better than the bivalent classical logic.

Jayatilleke also points out that most examples illustrating the catuṣkoṭi in the Pāli Canon are of the form:

(1) S is P

(2) S is non-P

(3) S is P and S is non-P

(4) S is neither P nor non-P[105]

Let p be the proposition "S is P." Then these four states can be rewritten as:

(1) p is true

(2) p is false

(3) p is both true and false

(4) p is neither true nor false

The relationship to the four truth-values of FDE is quite obvious. Additionally, catuṣkoṭi literally means "four corners", which evokes the diamond-shaped lattice used to visualize the four truth-values of FDE.

There has been some disagreement about which four-valued formal logic best models the Buddhist catuṣkoṭi: Cotnoir[106] for instance argues that FDE is not well-suited for it. These discussions are far beyond the scope of this thesis; what this section hopes to achieve is merely to provide another way of making sense of the four truth-values of FDE and to point out that there is no reason why a bivalent logic such as classical logic is more intuitive than a four-valued one.

Section 3 – The semantics of FDE

The logic of FDE can be interpreted in four different ways. I will present the four-valued semantics first, as it is the easiest way to give the truth-tables, although the full justification for these tables will be given by the two following semantics: the two-valued Dunn semantics and the algebraic semantics. Finally, the Routley star semantics will be given.

[105] Jayatilleke, K. N. "The Logic of Four Alternatives," 78
[106] Cotnoir, "Nagarjuna's Logic"

3.1 The four-valued semantics

In the four-valued semantics, the set of truth-values $\mathcal{V}_4$ is the set $\{t, f, i, j\}$. There is an interpretation v which maps formulae onto truth-values in $\mathcal{V}_4$. The interpretation v is a truth-function: $v(p)$ takes one value among t, f, i and j. T.J. Smiley[107] provided the "characteristic matrices", or truth-tables, of the logic FDE. Most of these values can be found by taking the truth-tables from classical logic, K_3 and LP; the only new cases are the ones where the inputs are i and j. The full justifications for these truth-tables will be given with the following two semantics.

Negation ¬

¬	
t	f
i	i
j	j
f	t

As in classical logic, the negation of 'true' is 'false', and vice-versa. The intermediate values each take their own value as their negation, just as they do in K_3 and LP.

Conjunction ∧

∧	t	i	j	f
t	t	i	j	f
i	i	i	$\boldsymbol{f}$	f
j	j	$\boldsymbol{f}$	j	f
f	f	f	f	f

These entries are derived from the following principle: all the values from classical logic, K_3 and LP are kept. There are only two new entries, in bold script, corresponding to the case where one of the conjuncts takes the value i and the other the value j. The reason why those two entries

[107] TJ Smiley pointed this out in correspondence, see Belnap, "A useful four-valued logic", 16, and Anderson and Belnap, *Entailment: The Logic of Relevance*, 161.
The truth-tables can be found for instance in Priest, *An Introduction to Non-Classical Logic: From If to Is*, 146.

take the value f is evident with the following two semantics, in particular with the algebraic semantics. For now a justification can be given with the intuitive understanding of i as 'neither true nor false' and j as 'both true and false'. Recall the principles of conjunction: the conjunction is only true if both conjuncts are true, and the conjunction is false whenever one of the conjuncts is false. Let $v(p) = i$ and $v(q) = j$. It is not the case that both conjuncts are true, so $v(p \wedge q) \neq t$. If we think of j as 'both', however, q is both true and false, so one of the conjuncts is false. Hence, $v(p \wedge q) = f$.

Disjunction ∨

∨	t	i	j	f
t	t	t	t	t
i	t	i	$\boldsymbol{t}$	i
j	t	$\boldsymbol{t}$	j	j
f	t	i	j	f

As with conjunction, most of these entries can be taken directly from classical logic, K_3 and LP. The only two new entries, in bold script, are the case where one of the disjuncts takes the value i and the other the value j. It will be clearer why those entries take the value t with the other semantics, but interpreting i as 'neither true nor false' and j as 'both true and false' provides some justification. Recall that a disjunction is only false when both disjuncts are false, and true wherever one of the disjuncts is true. Let $v(p) = i$ and $v(q) = j$. It is not the case that both disjuncts are false, so $v(p \vee q) \neq f$. If we think of j as 'both', however, q is both true and false, so one of the disjuncts is true. Hence, $v(p \vee q) = t$.

3.2 The two-valued Dunn semantics

In the two-valued semantics due to Michael Dunn[108], the set of truth-values $\mathcal{V}_2$ is the same as in classical logic, namely the set $\{t, f\}$. Nevertheless, the intuitive interpretation 'true, false, both, neither' will become quite clear.

The interpretation which maps formulae onto truth-values in $\mathcal{V}$ is no longer a function but a relation: the main difference is that a relation can have several outputs for one input, and not

[108] Dunn, "Intuitive semantics for first-degree entailment and "coupled trees""

just one as for functions. A proposition p may take the value t, f, or take both t and f or neither of them.

Let us denote the truth-relation with the symbol $\mathcal{R}$. To write that the proposition p is true, we write $p\mathcal{R}t$; to write that the proposition p is false, we write $p\mathcal{R}f$. But the proposition can also be in relation to both t and f (we can have $p\mathcal{R}t$ and $p\mathcal{R}f$) or not be in relation to anything at all. In other words, it is quite natural to interpret this as meaning that a proposition p can be true, false, both true and false, and neither true nor false.

The connectives ¬, ∧ and ∨ are defined as follows[109]:

Negation ¬	$\neg p\mathcal{R}t$ iff $p\mathcal{R}f$
	$\neg p\mathcal{R}f$ iff $p\mathcal{R}t$
Conjunction ∧	$(p \wedge q)\mathcal{R}t$ iff $p\mathcal{R}t$ and $q\mathcal{R}t$
	$(p \wedge q)\mathcal{R}f$ iff $p\mathcal{R}f$ or $q\mathcal{R}f$
Disjunction ∨	$(p \vee q)\mathcal{R}t$ iff $p\mathcal{R}t$ or $q\mathcal{R}t$
	$(p \vee q)\mathcal{R}f$ iff $p\mathcal{R}f$ and $q\mathcal{R}f$

The new entries in the truth-tables of the previous semantic can be found by applying these rules. If $v(p) = i$, then we have neither $p\mathcal{R}t$ nor $p\mathcal{R}f$. If $v(q) = j$, we have both $q\mathcal{R}t$ and $q\mathcal{R}f$. Consequently, for conjunction, since we do not have both $p\mathcal{R}t$ and $q\mathcal{R}t$, we do not have $(p \wedge q)\mathcal{R}t$. However, since $q\mathcal{R}f$, we have $(p \wedge q)\mathcal{R}f$. This is why $v(p \wedge q) = f$.

For disjunction we do not have both $p\mathcal{R}f$ and $q\mathcal{R}f$, so it is not the case that $(p \vee q)\mathcal{R}f$. However, $q\mathcal{R}t$, and therefore $(p \vee q)\mathcal{R}t$. Hence, $v(p \vee q) = t$.

There is another connection between the two-valued semantics and the four-valued one. The set of truth-values in the two-valued semantics is the set $\mathcal{V}_2 = \{t, f\}$. The power set of a set contains all possible subsets of that set. The power set of $\mathcal{V}_2$ is $\mathcal{P}(\mathcal{V}_2) = \{\{t\}, \{f\}, \emptyset, \{t, f\}\}$. Identifying the empty set $\emptyset$ with 'neither t nor f' and therefore with i, and the subset $\{t, f\}$ with 'both t and f' and therefore with j, we can make a one-to-one correspondence between $\mathcal{P}(\mathcal{V}_2) = \{\{t\}, \{f\}, \emptyset, \{t, f\}\}$ and $\mathcal{V}_4 = \{t, f, i, j\}$.

[109] See Dunn, "Intuitive semantics for first-degree entailment and "coupled trees,"" 156; or 1976 p.156 or Priest, *An Introduction to Non-Classical Logic: From If to Is*, 143

3.3 The algebraic semantics

The logic FDE also forms a structure called a De Morgan algebra,[110] which is best presented by Makinson[111]. (Note, however, that he calls this logic the "logic of the de Morgan implication", and not 'FDE.' A more solid connection between FDE and the De Morgan algebra is presented by Font[112].

A De Morgan algebra is a structure comprised of a set, a unary operation, two binary operations and two special elements. The binary operations are often called $+$ and $\times$, and the two special elements 0 and 1, but for the sake of simplicity I will use the symbols of FDE.

The De Morgan algebra that emerges in the context of FDE is a structure made of the set of truth-values $\mathcal{V}_4 = \{t, f, i, j\}$ together with the unary operation $\neg$, the two binary operations $\wedge$ and $\vee$ and the two special elements t and f.

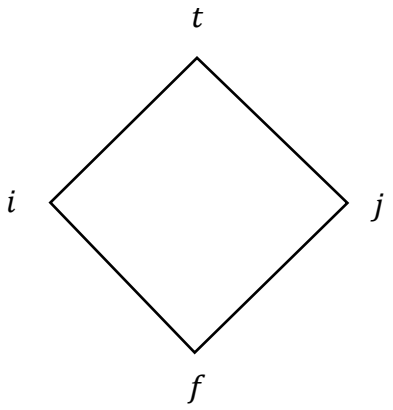
The interpretation will be the truth-function v.

The unary operation $\neg$ is a self-inverse function, i.e., $\neg\neg p = p$ that satisfies the De Morgan laws: $\neg(p \wedge q) = (\neg p \vee \neg q)$ and $\neg(p \vee q) = (\neg p \wedge \neg q)$.

The De Morgan algebra forms a bounded distributive lattice which, in the context of FDE, is easiest defined through a binary order relation $\leq$ on the set of truth-values $= \{t, f, i, j\}$. The relation $\leq$ is a partial order: it is reflexive ($a \leq a$), antisymmetric (if $a \leq b$ and $b \leq a$ then $a = b$) and transitive (if $a \leq b$ and $b \leq c$ then $a \leq c$). The set $\{t, f, i, j\}$ with the relation $\leq$ form a partial order (or poset). This means that not all elements of the set can be ordered. In the set $\{t, f, i, j\}$, there are two partial orders: $f \leq i \leq t$ and $f \leq j \leq t$.

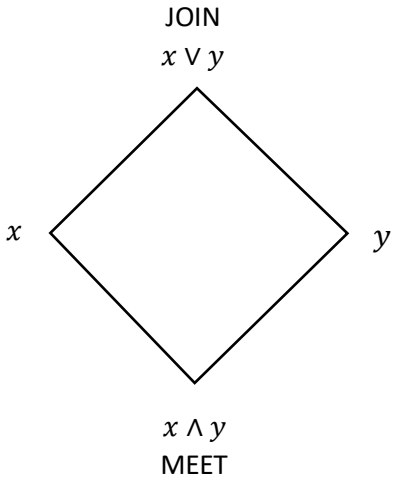
Partially ordered sets are easiest to represent with a Hasse diagram:

[110] History and background of De Morgan algebras is Béziau, "A History of Truth-Values," 280.
[111] Makinson, *Topics in Modern Logic*
[112] Font, "Belnap's Four-Valued Logic and De Morgan Lattices"



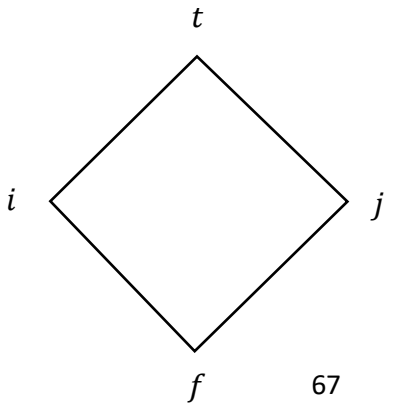
For any $x, y \in \{t, f, i, j\}$, $x \leq y$ iff x is lower on the diagram than y. That is, we have $i \leq t$, f$\leq t$ and $f \leq j$, but $t \not\leq i$ and $j \not\leq i$.

We define conjunction $\wedge$ and disjunction $\vee$ through the notion of 'join' and 'meet'. The join is the least upper bound and the meet is the greatest lower bound. Consider the following Hasse diagram:



Here the maximal element $x \vee y$ is the join of x and y and the minimal element $x \wedge y$ is the meet of x and y.

In order to find the value of $p \wedge q$ and $p \vee q$ on the diagram, we look for respectively the meet and join of $v(p)$ and $v(q)$.



If $v(p) = i$ and $v(q) = t$, their meet is the minimal element i, hence $v(p \wedge q) = i$. If $v(p) = j$ and $v(q) = f$, their meet is the minimal element f, so $v(p \wedge q) = f$. We can check that the entire truth-table of conjunction from section 3.1 can be found in this manner. In the case where $v(p) = i$ and $v(q) = j$, which was the only new case in FDE, $p \wedge q$ gets the value f since this is the minimal element of i and j.

The same goes for the disjunction $p \vee q$. If $v(p) = i$ and $v(q) = t$, their meet is the maximal element t, hence $v(p \vee q) = t$. If $v(p) = j$ and $v(q) = f$, their meet is the maximal element j, so $v(p \vee q) = j$. In the case where $v(p) = i$ and $v(q) = j$, the maximal element on the diagram is t, which is why $v(p \vee q) = t$.

This provides another justification for the truth-tables in 3.1.

3.4 The Routley star semantics

There is one last semantics for FDE developed by Val and Richard Routley (later Richard Sylvan and Val Plumwood)[113]. It involves a different understanding of negation: assume that each world w is associated to 'star world' w^*. Negation is defined as follows: $\neg p$ is true at w if p is false at w^*.

Formally, a Routley interpretation is a structure $\langle W, *, v \rangle$, where W is a set of worlds, $*$ is a function between worlds such that $w^{**} = w$, and v is an interpretation that assigns each formula the truth-value t or f, according to the following rules:

$v_w(p \wedge q) = t$ if $v_w(p) = t$ and $v_w(q) = t$; $v_w(p \wedge q) = f$ otherwise

$v_w(p \vee q) = t$ if $v_w(p) = t$ or $v_w(q) = t$; $v_w(p \vee q) = f$ otherwise

$v_w(\neg p) = t$ if $v_{w^*}(p) = f$; $v_w(\neg p) = f$ otherwise

The link between this semantic and the relational two-valued Dunn semantics is that the truth-relation $\mathcal{R}$ is equivalent to a pair of worlds w and w^*: $v_w(p) = t$ iff $p\mathcal{R}t$ and $v_{w^*}(p) = f$ iff $p\mathcal{R}f$.[114]

This interpretation is not very intuitive and is just given here as a matter of information.

[113] Routley and Routley, "The Semantics of First Degree Entailment." See also Priest, *An Introduction to Non-Classical Logic: From If to Is*, 151 and Omori and Wansing, "40 years of FDE: An Introductory Overview," 1024.
[114] Priest, *An Introduction to Non-Classical Logic: From If to Is*, 153; for proofs see 159-160.

Section 4 – Implication in FDE

The logic FDE famously has difficulties with implication. Priest writes that "FDE contains no real conditional connective"[115] and Omori and Wansing write that "FDE is known to lack a "decent" conditional."[116] This section will first present material conditional and mention some of its weaknesses before presenting other options to define implication. I will defend __

Note on the terminology: I use "implication" as a general term for all consequence relations and conditionals, and reserve "conditional" for the material conditional.

4.1 Material conditional

In classical logic, K_3 and LP, implication $\rightarrow$ is defined as material conditional $\supset$ through disjunction and negation: $p \supset q$ iff $\neg p \vee q$. If we define implication in that way, it gives us the following truth-table:

$v(\neg p \vee q)$		$v(q)$			
		t	i	j	f
$v(p)$	t	t	i	j	f
	i	t	i	t	i
	j	t	t	j	j
	f	t	t	t	t

The problem with material conditional is that it inherits the weaknesses of material implication in K_3 and LP: as we saw in chapter 4, the law of identity ($\vdash p \supset p$) is not valid in K_3, so it does not hold in FDE either. Modus ponens ($p, p \supset q \vdash q$) is not valid in LP, so neither is it in FDE.[117] Another issue is that transitivity ($p \supset q, q \supset r \vdash p \supset r$) is not valid in LP (take $v(p) = t$, $v(q) = j$, and $v(r) = f$) and is therefore not valid in FDE either.

As I will develop in section 5, the most important desideratum for this thesis is to have a T-schema. Material conditional in FDE is simply too weak: a T-schema based on it would not be sufficiently strong.

[115] Priest, "The logic of the catuskoti catuṣkoṭi," 34.
[116] Omori and Wansing. "40 years of FDE: An Introductory Overview," 1035
[117] Ibid.

Some choose nevertheless to keep material conditional as the main conditional in FDE; this is for instance the choice Graham Priest makes when developing the modal extension K_{FDE}.[118]

4.2 Implication from relevant logic

However, there are other ways to define an implication connective. The logic FDE was developed in the context of relevant logic, where one of the goals was to determine when $p \rightarrow q$ is valid based on relevance concerns: we want the antecedent p to be somehow *relevant* to the consequence q (see section 2.2 in this chapter). In this subsection I will present the implication developed for FDE for the purpose of relevant logic. Both Anderson and Belnap and Makinson develop an implication based on relevant logic; I will show that they are equivalent. At the end, I will show how this implication satisfies the variable sharing principle.

For reason of simplicity, it is easier to investigate relevance when there is only one implication connective in the sentence, that is, when the sentences (if they contain an implication at all) are of the form $A \rightarrow B$, where A and B do not contain an implication. The word 'degree' in 'first-degree entailment' refers to the degree of nesting of arrows[119]. A *zero-degree formula* is a formula that contains no implication, a *first-degree formula* (also called *first-degree entailment*) is a formula containing only one implication, and so on. Sentences in FDE can at most have one implication $A \rightarrow B$, where the formulas A and B must be implication-free, and can only contain disjunction, conjunction and negation.

With the de Morgan laws, any disjunction in a formula can be restated with the help of conjunction and negation, and any conjunction can be restated with disjunction and negation. If a formula has been restated using only disjunction and negation, we say that it is in *disjunctive normal form*. If a formula has been restated using only conjunction and negation, we say that it is in *conjunctive normal form*. We say that $A \rightarrow B$ is in *normal form* when A is in disjunctive normal form and B is in conjunctive normal form.[120]

Anderson and Belnap thus define an implication based on concerns from relevant logic. For instance, they do not want the law of excluded middle or the principle of explosion to be valid, but they want the law of identity to hold.[121] It would be beyond the scope of this thesis to go more in detail into Anderson and Belnap's argumentation; I will simply give here the truth-

[118] Priest, "Many-valued modal logics: a simple approach."
[119] Anderson and Belnap, *Entailment: The Logic of Relevance and Neccessity*, 151
[120] Omori and Wansing. "40 years of FDE: An Introductory Overview," 1022
[121] A list of desirable and undesirable examples can be found in Anderson and Belnap, *Entailment: The Logic of Relevance and Neccessity*, 154

table of the implication they establish on the grounds of criteria from relevant logic[122]:Note that all the outputs are classical.

$\rightarrow$	t	i	j	f
t	t	f	f	f
i	t	t	f	f
j	t	f	t	f
f	t	t	t	t

This corresponds to the implication Makinson[123] defines through the algebraic semantics: $p \rightarrow q$ is *valid* iff $\nu(p) \leq \nu(q)$ for any assignment of values. Recall that in the algebraic semantics, the truth values $\{t, f, i, j\}$ form two partial orders: $f \leq i \leq t$ and $f \leq j \leq t$. The easiest is to visualise this on the diamond lattice of the truth-values of FDE.

Using this criterion, we can then check for which values of $\nu(p)$ and $\nu(q)$ we have $\nu(p) \leq \nu(q)$ and establish a truth-table. We write a check mark ✓ for valid and an x mark ✗ for invalid.

$\nu(p) \leq \nu(q)$		$\nu(q)$			
		t	i	j	f
$\nu(p)$	t	✓	✗	✗	✗
	i	✓	✓	✗	✗
	j	✓	✗	✓	✗
	f	✓	✓	✓	✓

Replacing the check-mark ✓ with t and the x mark ✗ with f, this truth-table is identical to the one established by Anderson and Belnap.

With implication defined in this way, the law of identity holds: $\nu(p) \leq \nu(p)$ for any p.

Modus ponens also holds:

p	q	p	$p \rightarrow q$	q

[122] Anderson and Belnap, *Entailment: The Logic of Relevance and Neccessity*, 162
[123] Makinson, *Topics in Modern Logic*, 33

t	t	t	✓	t
t	i	t	✗	i
t	j	t	✗	j
t	f	t	✗	f
i	t	i	✓	t
i	i	i	✓	i
i	j	i	✗	j
i	f	i	✗	f
j	t	j	✓	t
j	i	j	✗	i
j	j	j	✓	j
j	f	j	✗	f
f	t	f	✓	t
f	i	f	✓	i
f	j	f	✓	j
f	f	f	✓	f

There is no case where p is designated, $p \to q$ is valid, and q is undesignated, so modus ponens holds.

Finally, we can check that this implication satisfies the variable sharing principle, which was the goal of relevant logic. To do so we check that if α and β have no propositional letters in common, there can be no tautology of the form $\alpha \to \beta$. This proof is taken from Makinson.[124]

Assign i to all propositional letters in α $(\alpha_1, \alpha_2, \dots, \alpha_n)$ and j to all propositional letters in β $(\beta_1, \beta_2, \dots, \beta_m)$ for some $n, m \in \mathbb{N}$.

Then for any $x, y \in \mathbb{N}$, $\quad v(\alpha_x) = i$

$v(\neg\alpha_x) = i$

[124] Makinson, *Topics in Modern Logic*, exercice 48, p.33 and p.94

$$v(\alpha_x \wedge \alpha_y) = i$$
$$v(\alpha_x \vee \alpha_y) = i$$

By induction, no operation involving the propositional letters in α will change its truth-value to anything but i. Thus $v(\alpha) = i$.

By a similar argument, $v(\beta) = j$.

Since $i \not\leq j$, we do not have $\alpha \rightarrow \beta$ if α and β have no propositional letters in common.

Thus, relevant logic provides us with an implication which satisfies the variable sharing principle, and also the law of identity and modus ponens. The only possible criticism is that the outputs of this implication are the two classical truth-values, and as such it may appear to some as being too strong, and not sufficiently in the spirit of FDE.

4.3 Other conditional operators

Relevant logic provides an implication for formulas in normal form, but there is still an ongoing debate about the conditional connective. The logic FDE can be extended in other ways so as to include an implication.

Omori and Wansing give several options[125], of which I will only cite the best-known, $\rightarrow_{excl}$. This is purely for the sake of curiosity; I will not use this implication further. Like material conditional, $\rightarrow_{excl}$ is defined through negation and disjunction, but it uses the so-called 'exclusion negation', which I will present shortly, instead of the classical negation. In other words, $p \rightarrow_{excl} q$ is defined as $\neg_{excl} p \vee q$.

The exclusion negation $\neg_{excl}$ (also called presupposition-denying negation) maps 'true' to 'false' and 'false' to 'true', like classical negation, but it also maps the negation of a proposition lacking a classical truth-value to 'true'.[126] For instance, if we consider that the sentence "I have a drawing of a round square" lacks a truth-value, since 'round square' fails to denote, then the sentence "I do not$_{\text{excl}}$ have a drawing of a round square" is true. In the context of FDE, the exclusion negation will map the undesignated intermediate value i to 'true' t and the designated intermediate value j to 'false' f. This can be understood intuitively by considering i as 'neither'

[125] Omori and Wansing. "40 years of FDE: An Introductory Overview," 1036
[126] Beaver, Geurts, and Denlinger, "Presupposition"

and j as false. The exclusion negation maps anything that is true to f, which is why t and j map to f. The values f and i have nothing to do with t so when negated they map to t.

Exclusion negation $\neg_{excl}$

$\neg_{excl}$	
t	f
i	t
j	f
f	t

We can then determine the truth-table for $\neg_{excl} p \vee q$, that is, $p \rightarrow_{excl} q$.

$v(\neg_{excl} p \vee q)$		$v(q)$			
		t	i	j	f
$v(p)$	t	t	i	j	f
	i	t	t	t	t
	j	t	i	j	f
	f	t	t	t	t

As seen in the truth-table, if the antecedent is not true (i or f), then the conditional becomes true. If not (if the antecedent is t or j), the conditional takes the truth-value of the consequent.

There is still an active debate going about which conditional is best for FDE, and surveying them would be far beyond the scope of this thesis, so I will not take position any further. It is sufficient to know that there are multiple options, of which a few have been presented.

Section 5 – The T-schema in FDE

One of the overarching goals of this thesis is to have a truth predicate defined through the T-schema. However, as seen in chapter 3, there are several ways to formalise it, either through a biconditional $\leftrightarrow$ or through a double entailment $\dashv\vdash$ or two $\vDash$. What the T-schema expresses is an equivalence between a proposition and truth being predicated of that proposition.

The first option is to define the T-schema using the meta-language: this is the option chosen by Priest (subsection 5.1). However, this does not guarantee that both sides of the bi-entailment have the same truth-value, only that they have the same designated or undesignated status.

The second option is to formalise the T-schema with a biconditional, as $T(p) \leftrightarrow p$. However, the difficulties surrounding having an implication for FDE are transferred to the T-schema. Nevertheless, I will show that by using a biconditional based on the implication from relevant logic, we can get a satisfactory T-schema (subsection 5.2).

5.1 Priest's T-schema using bi-entailment

Instead of taking this biconditional as two implications (in which case the choice of implication becomes significant), another option is to consider the biconditional in the T-schema to be bi-entailment.

This is the strategy followed by Priest[127]. Here, the T-schema is defined as $T\langle A\rangle \dashv\models A$. Validity $\models$ is defined in the usual way as designation-preserving: $A \models B$ iff if $A \in \mathcal{D}$, then $B \in \mathcal{D}$.

This ensures that both sides of the bi-entailment have the same designated or undesignated status, however it does not guarantee the exact same truth-value. For this reason, I prefer defining the T-schema through a biconditional based on the implication through relevant logic.

5.2 Using the implication from relevant logic

Using the implication established by Anderson and Belnap (which coincides with Makinson's validity), we can establish biconditional $p \leftrightarrow q$ as $(p \rightarrow q) \wedge (q \rightarrow p)$ which gives us the following truth-table:

$\leftrightarrow$	t	i	j	f
t	t	f	f	f
i	f	t	f	f
j	f	f	t	f
f	f	f	f	t

Note that the output is only true following the diagonal, when both sides of the biconditional have the same value. In other words, the implication from relevant logic yields a biconditional

[127] Priest, Graham. "The logic of the catuṣkoṭi," 34

that is only true when both sides have the same truth-value: $\nu(p \leftrightarrow q) = t$ iff $\nu(p) = \nu(q)$. This is precisely what we need to establish a T-schema.

Writing, as earlier, T for the truth predicate, $\langle A \rangle$ for the name of the sentence A and $T\langle A \rangle$ for 'A is true', we can state the T-schema in FDE: $T\langle A \rangle \leftrightarrow A$.

This may also be taken as a further argument in favour of the implication from relevant logic.[128]

Section 6 – Conclusion: a solution to the liar in FDE

Let us consider the liar sentence again: $\lambda \coloneqq \neg T\langle \lambda \rangle$. In chapter 4, we saw how the liar paradox is caused by the principle of excluded middle and the principle of explosion, both of which have been abandoned in FDE.

There are two possible truth-values that can be given to the liar sentence: either the undesignated intermediary value i or the designated intermediary value j. By taking $\nu(\lambda) = i$ or $\nu(\lambda) = j$, no paradox emerges and the liar sentence remains unproblematic.

These two possibilities correspond to a paracomplete and paraconsistent solution respectively, and it is possible to argue in favour of taking the truth-value of λ to be i or j based on their intuitive interpretations as 'neither true nor false' or 'both true and false'. As we saw in chapter 5, there might be a preference for assigning 'both' to the liar sentence, as an argument can be made for its falsity but also for its truth. Formally however, the only difference between i and j is that one is designated and the other is not, so whether one wishes the liar sentence to have a truth-value tracked by logical entailment depends more on what exactly we want the role of logical entailment to be.

The advantage of FDE over a pluralism between K_3 and LP is not only that FDE provides a unifying framework, but that FDE can easily be strengthened into K_3, LP or classical logic. In chapter 5, I argued that pluralism was problematic because K_3 and LP are then considered equally as valid, which means that their logical principles are considered equally as fundamental. The issue with this is that in K_3, the principle of explosion is considered to be a fundamental law of logic whereas the principle of excluded middle is discarded, whereas in LP, it is the principle of excluded middle that is considered as fundamental while the principle of explosion is abandoned. The problem with pluralism is reconciling these principles being seen

[128] Even though, surprisingly, this implication is not mentioned in the overview paper by Omori and Wansing.

as equally fundamental laws of logic on one hand while simultaneously being easily discarded on the other.

In FDE, neither of these two principles are considered to be fundamental laws of logic. Instead, they are merely additional principles that may be used to strengthen the logic in the right circumstances. Where there is no need for a designated intermediary value, for example outside of paradoxical situations, the logic can easily be strengthened with the principle of explosion into K_3. If there is no need for an undesignated intermediary value, for instance in a context where there are no future contingent statements or sentences with denotation issues, FDE can easily be strengthened into LP by adding the principle of excluded middle. Finally, if no intermediary values are needed at all, the two principles can be added so that the full strength of classical logic can be used.

The logic of FDE is therefore not that far removed from classical logic: in most cases classical logic can be used. Simply, FDE provides a framework for the puzzling and paradoxical cases that classical logic cannot handle. Finally, in a sense FDE is closer to classical logic than either K_3 or LP, as its truth values are directly related to the truth-values of classical logic.

Indeed, consider the power set of the classical truth-values. The power-set of a set is the set of all its subsets. With the classical truth-values $\{t, f\}$, four subsets can be formed: $\{t\}$, $\{f\}$, $\emptyset$ and $\{t, f\}$. These correspond to the four truth-values of FDE, with $\emptyset$ being related to 'neither true nor false' and thus to i and $\{t, f\}$ to 'both true and false' and thus to j.

Chapter 7 – The revenge of the liar

We saw in the last chapter that FDE provides a solution to the liar paradox. However, many solutions to the liar paradox fall prey to what is often called the "revenge of the liar," a slightly modified liar sentence which creates a new paradox. In this chapter I will construct a revenge sentence that FDE falls prey to, before exploring two possible responses to escape from this new paradox. First a dialethic solution, which embraces inconsistencies, and which involves constructing an infinity of truth-values, based on results by Roy Cook and Graham Priest (section 2). Second, a quietist solution, which argues that the revenge sentence should not have a traditional semantic value. Instead, the quietist will embrace a sub-logic ofFDE called FDE_e, with an additional truth-value standing for the 'ineffable' (section 3).

Section 1 – Introduction

Most solutions to the liar paradox are afflicted by another paradox called "the revenge of the liar." The revenge paradox is more of a phenomenon than an actual paradox: it is the collective name given to various paradoxes that emerge after a solution to the liar paradox has been proposed. JC Beall likens the liar and revenge paradoxes to a hydra[129]: as soon as a sword has been found to take the head off the liar paradox, another paradox immediately grows in its place.

In the family of paraconsistent and paracomplete solutions to the liar, the story of how the revenge paradox emerges goes as follows. A third truth-value of some type is introduced, and it is argued that the liar sentence should be assigned this truth-value. Next, the truth-values are put into two categories which I will call **true* and **false*. (For paracomplete solutions, **true* contains 'true' while **false* contains 'false' and the third truth-value. For paraconsistent solutions, **true* contains 'true' and the third truth-value, while **false* contains 'false.') Now consider the sentence "This sentence is **false*." In a similar way to the liar paradox, it is impossible to determine the truth-value of this sentence: a new paradox has reared its head. In the next paragraph, I will go through more precisely why a paradox emerges in FDE. It is not identical to the liar paradox, but the similarities are striking, which is why the revenge paradox is seen as the liar paradox slithering back.

[129] Beall, "Prolegomenon to future revenge," 4.

As FDE is both paracomplete and paraconsistent, it is unsurprisingly also subjected to the revenge phenomenon. Recall that FDE has four truth-values, t, f, i and j. These can be classified into two categories: designated and undesignated. The designated truth-values are t and j, while the undesignated truth-values are f and i. The categories **true* and **false* of the previous paragraph coincide with the designated and undesignated truth-values respectively.

Now, consider the sentence "This sentence is undesignated," which says of itself that it has the truth-vale i or f, and not the truth-value t or j. A paradox emerges when we try to assign a truth-value to the sentence. If the sentence has the truth-value i or f, then it is undesignated. In that case, the sentence which says of itself that it is undesignated is indeed undesignated, that is, the sentence is true (and has the truth-value t or possibly j). Conversely, if the sentence has the truth-value j or t, that is, if the sentence is designated, it is not undesignated, and what it says of itself is not true. In other words, the sentence has the truth-value f or i. To sum up, if the sentence has the truth-value i or f, then it is true (t or j); if the sentence has the truth-value j or t, then it is not true (i or f.) Although the logic FDE provides a solution to the liar paradox, it is vulnerable to the revenge phenomenon.

It would be far beyond the scope of this work to do an exhaustive analysis of all possible responses to the revenge sentence; instead, I will sketch out two-types of solutions.

Beall[130] classifies the responses to the revenge sentences into two categories: a "dialetheic" position and a "quietist" one. The dialetheic position (section 2) embraces inconsistencies: some sentences, such as the liar sentence, are both true and false, and others, such as the revenge sentence, are both designated and undesignated. What is required is a logical system which tolerates inconsistencies. The quietist position (section 3) argues that there is no point in giving a semantic value to such sentences: revenge sentences are neither designated nor undesignated, they do not fit into our categories and should not be made to conform.

Section 2 – The dialethic response

A dialethic type of response accepts inconsistencies. How would a dialethist respond to the revenge sentence in FDE? In this section I will first address why FDE, despite having a specific truth-value j constructed for the express purpose of dealing with inconsistencies in a dialethic fashion, cannot handle the paradoxical revenge sentence. Then I will suggest the framework for a new logical system, or rather a family of logics, that offers a solution to the revenge sentence

[130] Ibid., 4-5

as well as future revenge sentences, but which comes at the rather heavy price of accepting an infinity of truth-values. However, similar constructions can be found in the literature (Tarski, Cook and Priest), and I will argue that combining Cook's construction with Priest's justification makes the dialethic response more palatable.

2.1 FDE is not dialethic enough

The logic FDE should already be capable of handling inconsistencies: the principle of explosion does not hold, which means that an inconsistency does not automatically lead to deriving any arbitrary false statement. The original liar sentence is true if it is false and false if it is true, and can therefore be considered as 'both true and false,' and be assigned the intermediary truth-value j. Why can't the revenge sentence similarly be considered both true and false and simply be assigned this truth-value, since FDE has a pre-existing truth-value for that very purpose?

The issue with this, as noted in the first section, is that j is a designated truth-value: by assigning j to the revenge sentence, we are in effect claiming that a sentence defined as undesignated is designated. Although FDE has a truth-value available for inconsistencies, assigning j to the revenge sentence does not work as well as for the liar sentence.

Nevertheless, even without using the truth-value j for the revenge sentence, it is possible to claim that FDE can handle the revenge sentence and other inconsistencies simply in virtue of the principles of explosion and of excluded middle not holding. Without these two principles, sentences can indeed be both true and false, and this acceptance of inconsistencies should be applied more generally. A dialethist should simply embrace inconsistencies instead of shying away from them: the revenge sentence should be considered as both designated and undesignated, or perhaps as neither designated nor undesignated.

However, such a position may still appear unsatisfactory, as it is unable to assign a semantic value to the revenge sentence. Although FDE does not require that every sentence be assigned one and only one classical truth-value, it still has four clearly defined truth-values, and it is in a sense expected that meaningful statements should have a specific semantic value. This four-valued semantics constitutes what Priest refers to as an implicit "principle of the fifth-excluded."[131] This is why FDE is not in itself dialethic enough to handle the revenge sentence and will need to be tweaked.

[131] Priest, '*Quintum Non-Datur*', 17.

2.2 The dialethic solution

A second response is available for the dialethist. The dialethist dealt with the liar sentence by introducing a new truth-value j interpreted as 'both true and false.' Why not continue in the same spirit and introduce a new glutty truth-value, k, interpreted as 'both designated and undesignated?

This provides us with a paraconsistent-type solution to the revenge sentence. We know that if we assume that the revenge sentence is undesignated, we can derive that it is designated, and if we assume that it is designated, then it is undesignated. An argument can therefore be made for it being both designated and undesignated. Assign this new truth-value k to the revenge sentence, and the paradox is thwarted.

One immediate issue is how to make sense of a truth-value being both designated and undesignated. A truth-value is typically defined as designated if it is preserved by logical entailment, and undesignated if it is not. How could a truth-value be both preserved and not preserved by logical entailment? Alternatively, one might define logical entailment as a relation which preserves certain truth-values, namely the designated truth-values. However, this becomes very difficult with the presence of a truth-value that is both designated and undesignated. Should it be preserved by logical entailment? What is logical entailment in such a context? These are questions the dialetheic has to contend with.

A further issue is the immediate emergence of a new paradox. Our logic now has five truth-values: two designated values t and j, two undesignated values i and f, and finally the newly-minted k. I can now classify these truth-values into two new categories, which I will call 'hyperdesignated' and 'not-hyperdesignated.' The hyperdesignated truth-values are t, j and k, while the not-hyperdesignated values are i and f. Now consider the sentence "this sentence is not-hyperdesignated." Again, it is impossible to determine whether that sentence should have a hyperdesignated truth-value or a not-hyperdesignated truth-value: a new revenge paradox has emerged.

The dialethist pursuing this strategy is now facing a new problem, related to the issue of what designatedness and logical entailment mean in the presence of a truth-value that is both designated and undesignated. The new issue is what this new concept of 'hyperdesignatedness' may mean. There are two options for interpreting this term. The first is to argue that the division of the truth-values into 'hyperdesignated' and 'not hyperdesignated' is necessary in order to define logical entailment for this five-fold system. Logical entailment here is defined as

preserving the three 'hyperdesignated' truth-values. Because of the presence of k which is both designated and undesignated, it is no longer possible to define logical entailment as preserving designated values, but the concept 'hyperdesignated' allows us to fashion an entailment relation. The second option is to consider the category 'hyperdesignated' as nothing but the collection of the three truth-values t, j and k. In this case, the separation of the truth-values into the categories 'hyperdesignated' and 'not hyperdesignated' is nothing but a technical ploy to fashion the new revenge sentence "this sentence is not hyper-designated", which could just as well have been stated as "this sentence has neither the truth-value i nor f". However, this option does not say anything about what logical entailment should be.

I could also have introduced a new truth-value, l, interpreted as 'neither designated not undesignated,' for a paracomplete-type solution. There is some support for this kind of construction in the literature, namely the existence of a set of "antidesignated" truth-values, which does not have to be the complement of the set of designated values.[132] This opens a 'gap' between designated and undesignated truth-values, creating space for a gappy new truth-value l. In this case, the hyperdesignated truth-values would be t and j, while the not-hyperdesignated values would be i, f and l. This new truth-value could again be assigned as a solution to the revenge sentence. However, the same sentence "this sentence is not-hyperdesignated" emerges again as a new revenge sentence in this construction.

Alternatively, following the FDE framework, it is also possible to imagine a logic with both a new glutty truth-value k and a new gappy truth-value l. Here, there are three hyperdesignated values, t, j and k, and three not-hyperdesignated values f, i and l. In analogy with the use of paraconsistent logic for the liar sentence and paracomplete logic for the truth-teller sentence, as I argued for in chapter 5, it is possible to argue that the hyperdesignated value k should be applied to the revenge sentence "this sentence is undesignated" while the not-hyperdesignated value l should be applied to a 'designated-teller' sentence "this sentence is designated." However, such a six-valued logic also falls prey to the new revenge sentence "this sentence is not-hyperdesignated."

Nevertheless, the dialethist need not admit defeat immediately. The new revenge sentence can be thwarted with a new truth-value, interpreted either as 'both hyperdesignated and not-hyperdesignated' or as 'neither hyperdesignated nor not-hyperdesignated.' If we want to follow the FDE framework, we can introduce both values, which we can call m and n respectively.

[132] Shramko and Wansing, "Truth Values"

Now the revenge sentence "this sentence is not-hyperdesignated" can be assigned the truth-value m.

As expected, a new revenge paradox immediately appears. Let us introduce two new categories for our eight truth-values. The values t, j, k and m will be "meta-hyperdesignated" and the values f, i, l and n will be "not-meta-hyperdesignated." Now consider the sentence "this sentence is not-meta-hyperdesignated," which is yet another revenge sentence. It can be thwarted by introducing the new truth-values o and p, which will give rise to new categories, perhaps 'epi-meta-hyperdesignated,' which will lead to a new revenge sentence, and so on.

This can continue *ad infinitum*. Two new truth-values on the model of "both" and "neither" can always be introduced to counter the latest revenge sentence. All the truth-values can then be classified into our categories **true* (designated, hyperdesignated etc.) and **false* (undesignated, not-hyperdesignated etc.), with **true* containing t and all truth-values interpretable as 'both,' and with **false* containing those interpretable as 'neither.' Then a revenge sentence on the model "this sentence is **false*" emerges.

On the one hand, this shows the inevitability of the revenge phenomenon. A dialethist will always be confronted with a new revenge sentence, and no logical system modelled on FDE will manage to slay the revenge hydra while keeping the new heads from growing back. On the other hand, we have an infinite recipe that will always allow us to construct a solution to any revenge sentence of this type. Since there is an infinite succession of revenge sentences and solutions, whether it is the revenge sentences or the solutions that 'win' in the end becomes a matter of perspective.

Although this infinite construction can seem like a slow descent into truth-value madness, there are some elements in its defence.

2.3 Tarski's infinite layering

An infinite construction as a formal solution is nothing new in this philosophical tradition. Tarski's solution to the liar paradox involves an infinite hierarchy of languages and meta-languages, associated with an infinite layering of different truth predicates.[133] The infinite truth-value construction is not any worse than Tarski's infinite hierarchy: after all, the new truth-values are only needed to deal with revenge sentences that get progressively more bizarre and are unlikely to actually be used (how often does one need to express that a sentence is not meta-

[133] Tarski, "The semantic conception of truth".

hyperdesignated?) In contrast, I argued in chapter 2 that self-reference is fairly ubiquitous, and that sentences expressing their own falsity are difficult to avoid. If advocating for infinite truth-predicates is an acceptable solution for something as innocuous as expressing the truth-value of a sentence, then infinite truth-values that are only needed in a very specific case should likewise be tolerated.

2.4 Roy Cook's infinite truth-values

Roy Cook also argues for the existence of an infinite number of truth-values, although he provides a more general reason for this. Although embracing a construction with infinite truth-values can seem like a heavy price to pay, Cook argues for a different understanding of what a truth-value is to justify the existence of infinite truth-values. In his words, this makes the infinite truth-value construction "intuitive and roughly what we should have expected all along" instead of "shocking and obviously absurd."[134]

Cook centres his argument around semantic paradoxes like the liar paradox. These sentences are problematic as they cannot easily be assigned a classical truth-value. He then describes the typical non-classical solution which consists in introducing a new truth-value, which he calls "pathological." Cook then constructs the revenge sentence (which he calls 'strengthened liar') "this sentence is either false or pathological."[135] This corresponds to my sentence "This sentence is **false*."

However, it must be noted that this revenge sentence does not apply to paraconsistent solutions, even though Cook intends for it to work for any logic with a third truth-value, stating explicitly that he does not wish to take a stand in the paracomplete/paraconsistent debate. The value "pathological" is taken as being mutually exclusive with 'true' and 'false', whereas the paraconsistent third-value is typically interpreted as 'both true and false.' If 'pathological' is taken as leaning 'both true and false,' then the sentence "this sentence is either false or pathological" should be given the 'pathological' paraconsistent third truth-value, and would not trigger a paradox.

The way out of the revenge paradox is to introduce a new truth-value, "pathological$_2$", which will in turn give rise to a new revenge paradox: "This sentence is false or pathological or pathological$_2$." This construction can then continue infinitely.

[134] Cook, "What Is a Truth Value and How Many Are There?" 194
[135] Ibid., 190

Cook's construction is effectively very similar to the one I suggested, the only difference being that in the FDE context, two new truth-values are added at each stage, whereas Cook only adds one and provides a more general framework for all paracomplete logics.

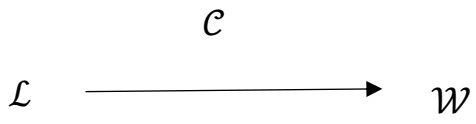
Although advocating for an infinity of truth-values may seem to be an absurd but technical solution which wilfully ignores what a truth-value is supposed to be, Cook argues instead that if we shift our understanding of what truth-values are, the infinite construction becomes natural.

Traditionally, a truth-value is nothing more than some object, assigned to a statement, which represents the statement's semantic status.[136] By contrast, Cook emphasizes the connection between the truth-value of a statement and the relationship this statement has with the world. A statement has a certain truth-value because it stands in relation to the world a specific way. A statement is true "iff what it says is the case" and is false "iff what it says fails to be the case."[137] According to Cook, there is an infinity of truth-values because there is an infinity of ways statements can relate to the world. Features such as vagueness and semantic paradoxes are not odd and bothersome features to be explained, but rather give insight into how language and the world relate to each other. Imagine a border-line case of grains of sand and consider the sentence "this is a heap." This sentence neither is the case nor fails to be the case: the statement stands in relation to the world in a different way. Cases like this one show that there are more ways a statement can relate to the world than merely the two classical ones. One solution is to introduce "a third relation between language and the world,"[138] where the statement "matches up partially, but not completely, with how the world is." This is a case where third truth-value is introduced to deal with such sentences. Another solution, embraced by degree-theorists, is to replace the two classical relations with a continuum of degrees of partial match between language and the world. In such a case, we get an infinity of truth-values ordered between 'true' and 'false.'

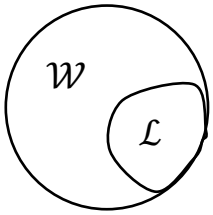
Both semantic paradoxes and vagueness explain why statements have an infinity of ways to relate to the world, which justifies postulating the existence of an infinity of truth-values. But Cook also gives a more general picture of how language and the world relate which explains why such an infinite construction is in fact natural and to be expected. In the remainder of this section I will explain Cook's theory as I picture it.

[136] Ibid., 183
[137] Ibid., 186
[138] Ibid., 188

The set-up starts as follows: a language $\mathcal{L}$ is in relation with a part of the world $\mathcal{W}$ through a class $\mathcal{C}$ of semantic relations.

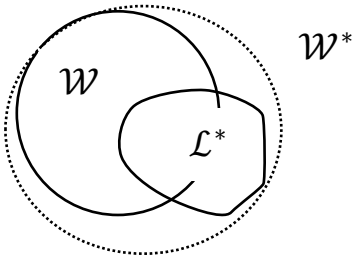


$\mathcal{W}$ is not intended to be the entire world or universe, but merely the part that is described by the language $\mathcal{L}$. This language is part of the world it describes, so a better illustration would in my opinion be the following:



However, the language $\mathcal{L}$ cannot adequately describe all the semantic relationships $\mathcal{C}$. This is because of semantic paradoxes such as the liar paradox, which according to Tarski (see chapter 2) make it impossible for any formal system to define its own truth predicate. In this context, it means that the language $\mathcal{L}$ does not contain a truth-predicate to express the truth or falsity of its own sentences. Therefore, we extend the language $\mathcal{L}$ to a richer language $\mathcal{L}^*$ in order to describe the semantic relations $\mathcal{C}$ satisfactorily.[139]

The old language $\mathcal{L}$ is not capable of describing all of $\mathcal{L}^*$, so $\mathcal{L}^*$ is not contained in the part of the world $\mathcal{W}$ described by $\mathcal{L}$. However, $\mathcal{L}^*$ is contained in the world $\mathcal{W}^*$ it describes.



[139] In order to argue for the extension from $\mathcal{L}$ to $\mathcal{L}^*$, Cook also cites Gödel's first incompleteness theorem, according to which any consistent formal system (as long as it is complex enough for some arithmetic (in particular it needs to have both addition and multiplication)) cannot prove all its true statements. Cook then jumps to the claim that the language $\mathcal{L}$ cannot express all true statements about itself, and thus for the need to extend to $\mathcal{L}^*$.

This construction continues infinitely, with the need to extend $\mathcal{L}^*$ into $\mathcal{L}^{**}$ in order to describe the relations $\mathcal{C}^*$ between $\mathcal{L}^*$ and $\mathcal{W}^*$ and so on. This is why the infinite construction with infinite truth-values becomes natural.

Although Cook provides a rationale for accepting a solution involving an infinite number of truth-values, one might not want to commit to this exact understanding of how language and the world relate. A different justification for having infinite truth-values can be found with the Dunn semantics.

2.5 The Dunn semantics and Priest's plurivalent logic

Although introducing new truth-values at will may seem arbitrary, the Dunn semantics[140] (see chapter 6 section 3.2) for FDE offers an alternative interpretation that may be more palatable. In the Dunn semantics, there are no additional truth-values beyond the two classical values true t and false f. The crucial difference with classical logic is that propositions are associated with truth-values via a truth-relation instead of a truth-function. With a truth-function, each proposition is associated with one and only one truth-value. With a truth-relation, however, a proposition can be in relation to one truth-value, several truth-values or none at all. If a proposition p is true, this is written $p\mathcal{R}t$, and if it is false, it is written $p\mathcal{R}f$. Instead of introducing the new truth-values i and j, the relational semantics writes '$\neg p\mathcal{R}t$ and $\neg p\mathcal{R}f$' to express that p is neither true nor false ($v(p) = i$), and '$p\mathcal{R}t$ and $p\mathcal{R}f$' to express that p is both true and false ($v(p) = j$).

The Dunn semantics provides a way to functionally get all four truth-values of FDE while still technically remaining in the classical two-valued realm. Using a truth-relation instead of a truth-function can therefore be seen as a technique to introduce new truth-values more covertly. Graham Priest has given a detailed analysis of this technique, which he calls "plurivalence."[141]

Let us consider the revenge sentence "this sentence is undesignated" in Dunn semantics. To express that a sentence p has a designated truth-value, it suffices to write $p\mathcal{R}t$; this covers both the case where p is only true (has the value t in the four-valued semantics) and the case where p is true and false (has the value j). To express that a sentence p does not have a designated truth-value, it suffices to write '$\neg p\mathcal{R}t$; this covers the case where p is only false (has the value f) and the case where p is neither true nor false (has the value i).

[140] Dunn, J. Michael. "Intuitive semantics for first-degree entailment and "coupled trees""

[141] Priest, Graham. "Plurivalent logics". Australasian Journal of Logic (11) 2014, Article no. 1. Pp. 2-13.

The revenge sentence in this semantics is thus '$\tau \coloneqq \neg\tau\mathcal{R}t$.' Although this looks like the liar sentence, the two-valued semantics would write '$\lambda \coloneqq (\lambda\mathcal{R}f \; and \; \neg\lambda\mathcal{R}t)$' to express the classical liar sentence. Let us attempt to relate τ to the truth-values.

(i) If $\tau\mathcal{R}t$ and $\neg\tau\mathcal{R}f$ (four-valued t), we get a contradiction, since by definition we have $\neg\tau\mathcal{R}t$. We cannot simultaneously have $\tau\mathcal{R}t$ and $\neg\tau\mathcal{R}t$.

(ii) If $\tau\mathcal{R}t$ and $\tau\mathcal{R}f$ (four-valued j), we get the same contradiction.

(iii) If $\neg\tau\mathcal{R}t$ and $\tau\mathcal{R}f$ (four-valued f), then τ claims something true (namely that $\neg\tau\mathcal{R}t$) and thus $\tau\mathcal{R}t$. The same contradiction $\tau\mathcal{R}t$ and $\neg\tau\mathcal{R}t$ arises.

(iv) If $\neg\tau\mathcal{R}t$ and $\neg\tau\mathcal{R}f$ (four-valued i), then the same reasoning as (iii) applies which leads to the usual contradiction.

In every possible case, the contradiction $\tau\mathcal{R}t$ and $\neg\tau\mathcal{R}t$ arises. The way out is therefore to accept the possibility that $\tau\mathcal{R}t$ and $\neg\tau\mathcal{R}t$. In other words, instead of merely accepting that a proposition p can be in relation to t and f or not, I also want the possibility that $p\mathcal{R}t$ and $\neg p\mathcal{R}t$ simultaneously. Further, if we accept that a proposition can both be and not be in relation to t at the same time, we can also open the possibility that a proposition be in relation to f and not be in relation to f at the same time as well, i.e., that $p\mathcal{R}f$ and $\neg p\mathcal{R}f$.

This allows for new possible relational cases, such as '$p\mathcal{R}t$ and $\neg p\mathcal{R}t$ and $p\mathcal{R}f$,' '$p\mathcal{R}t$ and $\neg p\mathcal{R}f$ and $p\mathcal{R}f$,' or indeed '$p\mathcal{R}t$ and $\neg p\mathcal{R}t$ and $\neg p\mathcal{R}f$ and $p\mathcal{R}f$.' In order to improve the readability of these cases, it is easier to see them as a relation on the four values t, j, i and f. The case where '$p\mathcal{R}j$ and $p\mathcal{R}f$' corresponds to '$p\mathcal{R}t$ and $\neg p\mathcal{R}t$ and $p\mathcal{R}f$', the case where '$p\mathcal{R}t$ and $p\mathcal{R}j$' corresponds to '$p\mathcal{R}t$ and $\neg p\mathcal{R}f$ and $p\mathcal{R}f$' and the case where '$p\mathcal{R}i$ and $p\mathcal{R}j$' corresponds to '$p\mathcal{R}t$ and $\neg p\mathcal{R}t$ and $\neg p\mathcal{R}f$ and $p\mathcal{R}f$.'

To sum up, the Dunn semantics allows us to get four different cases (corresponding to the four values of FDE) based on the two classical truth-values. By continuing to using a truth-relation instead of a truth-function, the same mechanism can be applied to these four values in order to get sixteen different cases, which can then be identified as sixteen new truth-values.

This process can then be repeated infinitely, providing a different way of constructing an infinite number of truth-values. The only difference with the dialethic method presented in the preceding subsection is that instead of going from n to $n + 2$ truth-values, the relational method goes from n to 2^n truth-values.

To conclude, we have one possible solution against the revenge paradox, by embracing a construction with an infinity of possible additional truth-values. Although such a logic is no longer the FDE logic, the way two new truth-values are added at each stage remains close to its 'spirit.' Both Cook and Priest offer justifications for accepting such a construction. Cook's theory is perhaps more commitment-heavy, as it relies on his understanding of how language and the world relate, whereas Priest's plurivalence technique only requires one to accept truth-relations instead of truth-functions.

Section 3 – The quietist response

The quietist response to the revenge sentences consists in claiming that such sentences have no semantic value. In the case of the revenge sentence for FDE; "this sentence is not undesignated," the quietist could claim that this sentence is neither designated nor undesignated and that it does not have a traditional semantic value. In this section I will introduce a new truth-value e which can be added to FDE in order to form the logic FDE_e. Although this value was at first introduced to deal with the 'ineffable,' it can be interpreted in other ways and can be assigned to the revenge sentence.

3.1 The revenge sentence is not well-formed

The first option for the quietist is to claim that there is something wrong with the revenge sentence itself. There are two possible issues with the revenge sentence. The first is the presence of self-reference. However, I argued in chapter 2 that self-reference was fairly ubiquitous and was to be preserved, which is why I will not further challenge the presence of self-reference.

The second possible issue is the term "undesignated," which is typically defined as "not designated," and thus as "not being preserved by logical consequence." Although the precise definition of what logical consequence is can be contentious, there is no reason why the term "undesignated" should be considered as badly defined. Moreover, in the context of FDE, I have simply defined "undesignated" as "either having the truth-value i or the truth-value j." Claiming that "undesignated" is not well-formed would lead us to give up on fundamental principles of how language is used to create meaning and would be too high a price to pay.

3.2 A fifth 'truth-value' in Buddhist logic

Another option for the quietist is to claim that although the revenge sentence is well-formed, it does not have a traditional semantic value. The remainder of the chapter will explore this claim.

Graham Priest has argued that FDE is well suited to formalise a Buddhist logic (see section 2.3 of chapter 6). In this logic, there are four possible semantic values (or 'four corners', called the 'catuṣkoṭi'): true, false, both and neither, which correspond to the four truth-values in FDE.

However, there seems to be an implicit fifth 'truth-value' in Buddhist logic. Priest[142] points out that in several texts, the Buddha refuses to choose among the four options in order to answer certain difficult questions. This is not merely because the question is not of interest or because of lack of knowledge; rather, neither of the four possible options are correct. When asked what happens to a fire that has gone extinct ("In which direction has the fire gone, -east, or west, or north, or south?"[143]), the Buddha simply answers "the question would not fit the case."[144] The statement "the extinct fire has gone north" is neither true, false, both true and false or neither; it is not syntactically wrong nor is it a mere epistemological issue. Neither of the four corners of the catuṣkoṭi can be assigned to the statement.

The four options are also explicitly denied in the *Mūlamadhyamakakārikā*[145]*:*

> Having passed into nirvāna, the Victorious Conqueror
>
> Is neither said to be existent
>
> Nor said to be nonexistent.
>
> Neither both nor neither are said. (MMK XXV: 17)

It is not possible to assign one of the four options of the catuṣkoṭi to whatever happens to the "Victorious Conqueror" after death. Even though Buddhist logic supposedly has only four semantic values, none of them is appropriate in this case.

In other passages as well, the four possible truth-values are explicitly mentioned, only to be denied. It is to formalise this that Graham Priest suggests an adaptation of FDE with a fifth truth-value, although one that is meant to represent the 'ineffable,' and which does not behave quite like a traditional truth-value.

[142] Priest, Graham. 'None of the Above: The Catuṣkoṭi in Indian Buddhist Logic'
[143] Ibid., 521; translation Radhakrishnan from and Moore, *A Source Book in Indian Philosophy*, 290.
[144] Ibid.
[145] Garfield, "*The Fundamental Principles of the Middle Way: Nāgārjuna's Mūlamadhyamakakārikā*."

3.3 The logic FDE_e

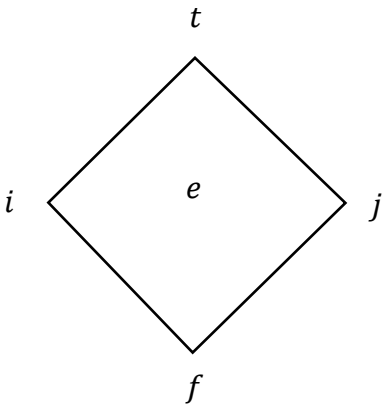
The logic FDE_e is a proper sub-logic of FDE with an added truth-value e to represent the ineffable (the value is called 'e' as it originally stood for 'empty'[146]). The designated values remain the same as in FDE, that is $\mathcal{D} = \{t, j\}$. For Priest, this newest truth-value should not be designated, since "neither they [the statements with the value e] nor their negations should be accepted"[147]. However, e must not be undesignated either (taking 'undesignated' to be different from merely not being designated), otherwise we run straight into the same type of revenge sentence that the dialetheic solution encountered. Thus, this new truth-value e is neither designated nor undesignated. Although this appears to be very similar to one of the dialethic solutions, there is a decisive difference between the value e and the dialethic value l 'neither designated not undesignated.'

Indeed, the value e is 'infectious': if any part of the input has the value e, then the output has the value e as well. In that sense Priest's e behaves a little like the intermediary value in Weak Kleene or in Bochvar's nonsense or meaningless logic (see chapter 5). For that reason, Garfield and Priest call it a "*sink*":[148] once any part of a statement is assigned e then the entire statement takes this value.

This also ensures that the two values i and e are different, even though they are both technically neither true nor false. Consider the disjunction "cats are mammals or there will be rain in Oslo in 76 days." If we accept that the truth-value of "there will be rain in Oslo in 76 days" is i (assigning i to future contingent statements), then the entire statement "cats are mammals and there will be rain in Oslo in 76 days" will have the value t, since it is always true that cats are mammals. But consider a proposition p which takes the truth-value e. Then the statement "cats are mammals or p" is no longer true, even though it is always the case that cats are mammals, and the statement takes the new truth-value e. The entire sentence, because a component has the value e, must also take this value.

This new value is typically[149] represented visually within the classic FDE diamond lattice:

[146] Garfield and Priest, "Mountains are just mountains."
[147] Priest, *the logic of the catuṣkoṭi*, p.37
[148] Garfield and Priest, "Mountains are just mountains," 77
[149] Ibid.



However, the value e could also have been placed anywhere outside of the diamond. As Garfield and Priest, who seem to be the first ones to use this diagram,[150] write, the value e is "an isolated point"[151] and is "incomparable with the other four values." Consider the central placement to be mostly an aesthetic choice (as well as an indication that the new value is equidistant from all the others); the main point is that contrary to the four other truth-values, the value e is not part of the partially ordered set represented by the Hasse diagram.

Some rules of FDE must be adapted, in particular the rule for disjunction-introduction,[152] but I will not go into all the details of FDE_e here. What is relevant is that it is a logic with a value e for the ineffable in addition to the four truth-values of FDE. The soundness and completeness FDE_e have been proved by Priest[153] and the full truth-tables have also been published.[154] Note finally that in earlier sources Priest calls the logic FDE_φ,[155] but in more recent seems to prefer FDE_e.[156]

3.4 The value e and the ineffable

Although the logic FDE_e with its additional truth-value e was constructed to formalise the 'ineffable' in Buddhist logic, other philosophical traditions also have a long tradition of dealing with the ineffable, and FDE_e can be applied far beyond Buddhist logic.

The concept of the ineffable has suffered from being associated with mysticism, but Silvia Jonas argues that it does not have to be "a horror for the coolly detached mind of the analytical

[150] Garfield and Priest, "Mountains are just mountains," 77
[151] Ibid.
[152] See e.g. Priest, 'None of the Above', 522 or Priest, "The logic of the catuskoti", 37-39
[153] Appendix of Priest, "The logic of the Catuskoti"
[154] Priest, "Natural Deduction Systems for Logics in the *FDE* Family."
[155] Priest, 'None of the Above' and "The logic of the catuskoti"
[156] For instance Priest. "Natural Deduction Systems for Logics in the *FDE* Family."

philosopher."[157] On the contrary, it appears throughout the history of philosophy, and it is rather odd that it has been largely ignored by analytical philosophy. Jonas' book not only analyses the concept but further offers a unified theory of the metaphysics of ineffability.

The ineffable is that which cannot be put into words, and as philosophy, especially academic philosophy, is a field which is almost exclusively done using words, it can seem incoherent to claim that the ineffable should be of philosophical interest at all. Nevertheless, it has been a subject of study for many philosophers and is difficult to avoid entirely; in Henry Sheffer's words "the spirit of ineffability in philosophy is subtly pervasive."[158]

Silvia Jonas offers an overview of the ineffable in the history of philosophy which I will not paraphrase. However, I would like to point out a few examples to illustrate how the topic of the ineffable emerges in very different contexts.

One of the foundational texts of Daoism is the *Dàodé Jīng*, written around 400BC and attributed to Laozi. The most important concept is that of the '*dào*' (道), which is notoriously impossible to translate and to define. It can be understood as 'way', 'path', road', 'guide', but also as 'method', 'manner', 'practice' and even 'speech'[159] [160]; in that regard it has often been compared to the Greek *logos*.[161] Most English versions now simply use 'the *dao*', as no translation could avoid taking too strong an interpretative position. Interestingly, the *dao* is not merely a concept that happens to be untranslatable from Ancient Chinese into other languages, it is also nearly impossible to define in Ancient Chinese. Indeed, the opening lines of the *Dàodé Jīng* concerns this very matter: 道可道，非常 道 (dào kě dào，fēi cháng dào), which means something along the lines of 'the *dao* that can be spoken of is not the (eternal/constant/actual) *dao*.'[162] The *dao* is something that cannot be expressed; if it could be expressed and talked about, it would no longer be the *dao*. The *dao* is in its very nature ineffable. Despite this, the entire text is nevertheless about the *dao*, attempting to give some understanding of what the *dao* is through the text.

The ineffable is also to be found in Plotinus, who argues for the existence of three fundamental metaphysical principles: the 'One' (also called the 'Good', 'Unity', 'Supreme'), the Intellect

[157] Jonas, *Ineffability and its Metaphysics: The Unspeakable in Art, Religion, and Philosophy*, 1
[158] Sheffer, "Ineffable Philosophies," 129.
[159]Hansen, Chad, "Daoism
[160] Stefon, "dao". .
[161] See for instance Zhang, *The Tao and the Logos: Literary Hermeneutics, East and West*.
[162] Boisen offers eight different translations in Boisen, *Lao Tzu's Tao-Te-Ching : A parallel translation collection*.

and the Soul.[163] Of the three, the 'One' is the most fundamental, as the existence of the Soul depends on the Intellect, which in turn depends on the One. The One does not ontologically depend on anything but exists necessarily. However, the One is ineffable[164]: "The One is in truth beyond all statement […] we can give it no name because that would imply predication: we can but try to indicate, in our own feeble way, something concerning it."[165] Similarly to the *dao*, the One is the central metaphysical concept, yet it is ineffable. Nothing can be directly predicated of it, but it is possible to discuss how it relates to other metaphysical entities. However, as Silvia Jones writes, some knowledge can be gained about it through language: "by approximation, that is, through similes and metaphors."[166]

The last philosopher I want to mention is Wittgenstein, who is the philosopher most associated with the ineffable and the "limits of language" in the analytic tradition. In the *Tractatus* in particular, several elements are said to be beyond language. There are too many different interpretative traditions around Wittgenstein, which disagree on how Wittgenstein engages with the ineffable, and it would be far beyond the scope of this chapter to engage with this. Since I merely want to show the ubiquity of the ineffable in different philosophical traditions, I will simply follow Jonas, who enumerates the following elements that according to Wittgenstein are beyond language: (i) the "harmony between thought, language, and reality", (ii) "fundamental logical relations between propositions", (iii) "the limits of thought", that is, what cannot be thought, since if it could be expressed, it would be thinkable, and (iv) the "metaphysics of experience".

In particular, the *Tractatus* ends with an injunction to silence for a range of traditional questions in metaphysics, ethics and aesthetics, which cannot be expressed and are beyond language. This is contained in the very last the proposition: "Whereof one cannot speak, thereof one must be silent."[167] Nevertheless, the remainder of the text must not be dismissed as nonsense, or, at the very least, it is nonsense that must be taken seriously. Although there are things that cannot be expressed by language, they can be 'shown' through language. This is where the famous ladder metaphor from *Tractatus* comes in: "My propositions serve as elucidations in the following way: anyone who understands me eventually recognizes them as nonsensical, when he has used

[163] Gerson, "Plotinus"

[164] Jonas, *Ineffability and its Metaphysics*, 11-12
[165] Plotinus, *The Divine Mind, Being the Treatises of the Fifth Ennead*, Book V, Ch. 3, Passage 13.
[166] Jonas, *Ineffability and its Metaphysics*, 12
[167] Wittgenstein, *Tractatus Logico-Philosophicus* §7

them—as steps—to climb beyond them. (He must, so to speak, throw away the ladder after he has climbed up it.)"[168] All the preceding propositions are in a sense 'nonsense,' but are nonetheless necessary in order to gain understanding. Although someone who has gained this understanding can now throw them away, the understanding would not have been gained without them. This understanding could not have been expressed in a proposition, it is not something expressible, yet it has been 'shown' through the text.

I would like to point out a common element in the three examples I chose. For Laozi, Plotinus and Wittgenstein, the most important elements are ineffable, whether it be the *dao*, the One, or the central topics of philosophy. The ineffable cannot be directly said, yet all three agree that language can nevertheless convey something meaningful about the ineffable.

Finally, the ineffable also emerges in the mathematical domain, where it is much more difficult to dismiss as 'mystical.' One example is the (Zermelo-)König's paradox.[169] The real numbers $\mathbb{R}$ contain non-denumerably many elements. However, it can be proved that it is only possible to finitely define denumerably many of them. This means that there are reals that cannot be finitely defined. Since the reals $\mathbb{R}$ can be well-ordered, the reals that cannot be finitely defined must have a smallest member. Thus, the paradox emerges: there is a number which cannot, by definition, by finitely defined, yet is defined through the finite phrase 'the smallest real that cannot be finitely defined.' Here is a number that cannot be described, but can be referred to in just eight words: it cannot be expressed through language, but language can say something meaningful about it.

We have seen that the ineffable has been a topic of importance throughout the history of philosophy. Having a logic capable of handling the ineffable is therefore useful, far beyond wanting to formalize a few local Buddhist texts. The logic FDE_e may therefore have a much broader range of applicability than the one imagined by Priest.

3.5 Revenge and the ineffable

Should the quietist then claim that the revenge sentence ought to be assigned the value e 'ineffable'? Is the truth-value of the revenge sentence not merely difficult to ascertain, but ineffable?

[168] Ibid. §6.54
[169] Miriam Franchella, 'In the footsteps of Julius König's paradox'

Jonas explicitly addresses semantic paradoxes in her book[170], but concludes that she does not consider the truth-value of semantic paradoxes such as the liar paradox to be ineffable. However, her main goal is to examine the metaphysics of ineffability, and for this purpose, semantic paradoxes are simply not very useful. Jonas points out that both paraconsistent and paracomplete solutions to the liar paradox exist, which both assign an expressible truth-value (i or j) to the liar sentence. Arguing that the truth-value of the liar sentence is ineffable would therefore mean opposing the non-classical solutions to the liar, which would according to her be difficult and unlikely to succeed.

However, the case is a little different for the revenge sentence. Section 2 of this chapter did indeed suggest new truth-values that the revenge sentence could take (k or l), but these additional truth-values and the infinite construction of truth-values they entail are harder to accept than the now well-known paracomplete and paraconsistent truth-values 'neither' and 'both.'

The more traditional cases where the ineffable appears have indeed been in rather particular contexts, such as Plotinus' fundamental metaphysical principle of the One or Wittgenstein's limits of language. The revenge sentence may seem too prosaic to be deemed ineffable. After all, it is an easy sentence to state – why should its truth-value be ineffable?

However, recall that Priest introduced the truth-value e as 'none of the above,' that is, as a fifth option when neither of the four corners of the 'catuskoti' fits. In the *Mūlamadhyamakakārikā*[171], the "Victorious Conqueror" neither exist, nor is nonexistent, nor is both, nor is neither. None of the four options are possible, which is why the fifth value e is introduced. An analogy can be made for the revenge sentence: we know that none of the truth-values of FDE can be applied, therefore the truth-value must be e.

In this case, we can consider FDE_e more structurally: it has four 'traditional' truth-values, two designated, two undesignated, and an additional 'infectious' value e. Interpreting e as the ineffable is no longer necessary, and the truth-value can be assigned to a sentence by showing that the four other options are not possible, which can be easier than arguing that the sentence represents the ineffable.

[170] Jonas, *Ineffability and its Metaphysics*, 77-79
[171] Garfield, *The Fundamental Principles of the Middle Way: Nāgārjuna's Mūlamadhyamakakārikā*

3.6 Taking FDE_e beyond ineffability – a solution to the revenge paradox

We can indeed imagine other interpretations for the e of FDE_e. As I briefly mentioned, FDE_e was first introduced by Garfield and Priest to deal with the 'empty' in Zen Buddhism,[172] which is not quite the same as the ineffable. The 'emptiness' refers to a stage where one realises that objects, such as mountains, are "empty of inherent existence" and that this emptiness is identical to their existence. This refers to a famous story[173] in Zen Buddhism: at first, before studying Zen Buddhism, objects are as they appear: mountains are mountains. After studying Zen Buddhism a little, one realises that things do not really exist: there is no such thing as a mountain with an independent ontological existence. The world is, in a sense, empty. However, there is a third stage after studying Zen Buddhism even longer, where objects regain their existence: the mountain was in fact a mountain after all. However, the last stage is not the same as the first: in the last stage the student is aware of the ultimate emptiness of the world, the mountain exists, but to exist is the same as being empty of inherent existence.

This is not the place to discuss ontology in Zen Buddhism further, what is relevant here is that the structure of FDE_e and the value e is applicable beyond 'the ineffable.' Indeed, it is applicable beyond the different strands of Buddhism.

In chapter 5, I briefly mentioned that the third-value in various paracomplete logics have been interpreted in various ways: for instance as undefined, undecidable, indeterminate, unprovable, unknown, meaningless, nonsensical and nonsignificant. There are several different paracomplete logics, but what ultimately matters is how the truth-tables and rules differ between them, not how the third value should be interpreted. This was why I chose to remain as neutral as possible in interpreting the two additional values of FDE.

There is therefore solid precedent for interpreting a truth-value in multiple ways. Although the fifth value e in FDE_e was developed to deal with specific elements in Buddhist philosophy, it can be applied in other contexts, such as for the revenge sentence.

In conclusion, instead of only applying e to propositions that satisfy some sort of 'ineffable' or 'Zen-Buddhism-emptiness' criteria, it should also be used wherever the four values of FDE do

[172] Garfield and Priest, "Mountains are just mountains," 76

[173] First attributed to Master Qingyuan in the Compendium of the Five Lamps, see Garfield and Priest, "Mountains are just mountains," 71.

not fit. The quietist response to the revenge sentence in FDE_e is therefore simply to assign to it the new truth-value e.

Further, this value thwarts the revenge phenomenon in a way that the additional truth-values from the dialethic response (k and l) fail to do. I could attempt to formulate a new revenge sentence: "this sentence is neither undesignated nor has the value e." If I call this sentence φ, it can be reformulated as $\varphi \coloneqq \neg(v(\varphi) = f \vee v(\varphi) = i \vee v(\varphi) = e)$. Since the value e is specifically designed to act as a 'fail-safe,' so to speak, and infect the rest, the new revenge φ takes on the value e. There is no contradiction between φ being defined as not having the value e and being assigned the value e.

3.7 An epistemic interpretation of FDE_e

Independently of Priest's development of FDE_e based on Buddhist logic, a similar fifth value has been introduced by logicians in the context of non-deterministic semantics. Avron and Zamansky[174] suggest using the value $\perp$ for formulas without any logical value, which is closely related to the truth-value e. In the context of FDE, D'Agostino and Solares-Roja[175] are currently developing a non-deterministic semantics for a five-valued logic based on FDE, that is, FDE with the additional value $\perp$. In this case, FDE is interpreted epistemically, where the four truth-values correspond to an agent's information.

Belnap argued (see chapter 6 section 2.1) that FDE was well-suited to represent the four possible stages a computer can be in: the computer has been told that a statement is true, told that it is false, told that it is true and false, and told neither. The new value $\perp$ is meant to represent a state of full ignorance, where there is not even enough information to choose one of the four options, for instance while a computer is still running or processing information.

3.8 Remaining challenges for the quietist

Some challenges remain for the quietist. In particular, throwing a new fifth truth-value at the revenge sentence may seem just a little too convenient. Accepting the existence of a fifth truth-value, in particular to deal with the 'ineffable' or meaninglessness is one thing, but how does the quietist know which sentences should be assigned e? Could it become an easy way out of interesting philosophical conundrums?

[174] Avron and Zamansky, "Non-Deterministic Semantics for Logical Systems," remark 17 p.13.
[175] D'Agostino and Solares-Rojas. "Towards Tractable Approximations to Many-Valued Logics: the Case of First Degree Entailment."

Contradictions, paradoxes and inconsistencies are often seen as driving forces in philosophy. They can trigger a state of *aporia*, a feeling of wondrous puzzlement that leads us to inquire further about concepts and theories. Very often, an inconsistency can be an indication that something has gone wrong or that some premise ought to be re-examined. The quietist may risk assigning the value e too quickly. How can we know whether we have reached a state where none of the four values of FDE is appropriate, and where e should be assigned, or if the premises and reasoning should be reassessed?

Be that as it may, the value e is not the first truth-value that can be accused of being a means of evading the difficulty. The same charges could have been aimed at the paraconsistent third-value j 'both true and false.' However, this logic has been a part of philosophy for quite a few decades now, and it has not been overused. Only a few sentences are said to be truly paradoxical, and the logic of paradox LP cannot be said to have stymied philosophical activity. Similarly, there is no real reason to fear that the 'ineffable' value e will be utilized exaggeratedly.

3.9 Conclusion

Although FDE falls prey to the revenge paradox, there are a few solutions. The first one is the dialethic response, which allows for the possibility that propositions should be both designated and undesignated. Although this solution is no longer in the logic FDE, it remains closely related, especially as the infinity of additional truth-values are constructed following the FDE framework. However, the dialethic response may seem too commitment-heavy, as it requires either that one accepts Cook's theory on how language and the world relate, or that one follows Priest in considering truth-relations instead of truth-functions.

The second solution to the revenge paradox is the quietist response which uses FDE_e. Similarly to the dialethic response, this solution is strictly speaking not within FDE. However, FDE_e is a very closely related sub-logic of FDE. Although FDE_e originally came with some commitments to the existence of the 'ineffable,' the logic and in particular the fifth truth-value e can also be interpreted differently, for instance epistemically to denote a complete lack of information. By opening for other interpretations of e, FDE_e becomes more easily applicable, which makes the quietist solution less commitment-heavy.

Conclusion

In the first part of this thesis, I have argued that if we are to examine the concept of truth as it is used in ordinary language, we need to keep self-reference and the T-schema. In the second part, I have argued that the logic FDE is the best non-classical logic in which to formalise truth.

Although some doubts may remain, I hope to have at least convinced the reader that FDE is a rich and interesting logic to study, in particular due to its multiple interpretations. Although it originates from concerns about relevant logic, it has been applied to understanding computers, databases and even Buddhist logic.

The logic FDE is indeed weaker than classical logic or its paraconsistent and paracomplete cousins, K_3 and LP, but it does has an enviable flexibility. By demoting the law of excluded middle and the principle of explosion from fundamental laws of logic to mere rules, FDE can very easily be strengthened back into these logics. What remains for further study, however, is to establish more precisely the contexts in which these rules can be added back to FDE. I have argued that the principle of exluded middle can be used in all cases, except where there is a metaphysical indeterminacy or an epistemic ignorance, such as in the case of future contingents, and that the principle of explosion must only be given up in cases of inconsistencies and paradoxes. This means that in much of the time, we can use the full strength of classical logic. The logic FDE is in fact not that far removed from classical grounds. However, the exact conditions for when the principles can be used must be clarified.

I have also argued in favour of a specific implication operator, derived from relevant logic, however, it is far from being the commonly accepted implication for FDE, and a lot more work would have to be done in order to present a solid case for this implication to become standard.

Additionally, FDE is vulnerable to the revenge paradox, and although I have sketched out two possible responses, they both need to be developed further. Moreover, both responses involve leaving the FDE logic, either for an 'FDE-style' construction with infinite truth-values for the dialethic response, or for the shores of the sub-logic FDE_e with its 'ineffable' truth-value. Although both solutions remain close to FDE, they do constitute a weakness for the argument that FDE is the best logic for dealing with the liar paradox. The logic FDE is however in very good company, as most solutions to the liar paradox fall prey to a revenge paradox of some kind; in that regard FDE is not any worse than other solutions.

Finally, I must acknowledge that the weakest link of this thesis is the defence of the T-schema based on truth in natural language. However, at the time of writing, this was not a subject that has been sufficiently studied. The future results from the *Truth without Borders* project will hopefully contribute to clarify this matter and determine whether the T-schema is an essential component of truth.

Bibliography

Anderson, Alan Ross & Belnap, Nuel D. *Entailment: The Logic of Relevance and Neccessity*, Vol. I. Princeton University Press. 1975.

Appiah, Kwame Anthony. “Race, Culture, Identity: Misunderstood Connections.” In *Color Conscious: The political morality of race*, edited by Anthony K. Appiah and Amy Gutmann. Princeton, NJ: Princeton University Press, 1996.

Avron, Arnon and Zamansky, Anna. “Non-Deterministic Semantics for Logical Systems.” In: Gabbay, D., Guenthner, F. (eds) *Handbook of Philosophical Logic*, vol 16. Springer, Dordrecht. 2011

Beall, Jc. “Prolegomenon to future revenge.” In *Revenge of the Liar: New Essays on the Paradox*, edited by Jc Beall, 1-30. OUP, 2007.

Beall, Jc, Michael Glanzberg and David Ripley. *Formal theories of truth*. Oxford University Press, 2018.

Beall, Jc, Michael Glanzberg, and David Ripley. “Liar Paradox”. *The Stanford Encyclopedia of Philosophy* (Fall 2020 Edition). Edward N. Zalta (ed.). https://plato.stanford.edu/archives/fall2020/entries/liar-paradox/.

Beaver, David I., Bart Geurts, and Kristie Denlinger, "Presupposition", *The Stanford Encyclopedia of Philosophy* (Spring 2021 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/spr2021/entries/presupposition/>.

Belnap, Nuel D. “A useful four-valued logic.” In J. M. Dunn & G. Epstein (eds.), *Modern Uses of Multiple-Valued Logic*. D. Reidel. 1977.

Béziau, Jean-Yves. "A History of Truth-Values". In Gabbay, Dov M.; Pelletier, Francis Jeffry; Woods, John (eds.). *Logic: A History of its Central Concepts*. North Holland. 2012.

Blau, U. *Die dreiwertige Logik der Sprache: ihre Syntax, Semantik und Anwendung in der Sprachanalyse*, Berlin: de Gruyter. 1978.

Bochvar, D. A. “Об одном трехзначном исчислении и его применении к анализу парадоксов классического расширенного функционального исчисления” (“On a

three-valued logical calculus and its application to the analysis of contradictions"), Rec. Math. [Математический сборник] N.S., 4(46):2, 287–308. 1938.
The paper is in Russian, see D.A. Bochvar & Merrie Bergmann (1981), "On a three-valued logical calculus and its application to the analysis of the paradoxes of the classical extended functional calculus", History and Philosophy of Logic, 2:1-2, 87-112

Boisen, B. *Lao Tzu's Tao-Te-Ching : A parallel translation collection*, Gnomad publishing, Boston. 1996

Boolos, George, John P. Burgess, and Richard Jeffrey. *Computability and Logic (5th edition)*. Cambridge University Press, 2007.

Bourget, David and Chalmers, David. *Philosophers on Philosophy: The 2020 PhilPapers Survey*. November 1, 2021.
Accessible at <https://philpapers.org/archive/BOUPOP-3.pdf>

Cappelen, Herman. *Fixing Language: An Essay on Conceptual Engineering*. Oxford University Press, 2018.

Chalmers, David. "What is conceptual engineering and what should it be?" *Inquiry* (2020).

Chomsky, Noam. *Syntactic structures*. Mouton & Co., 1957.

Cook, Roy T. "What Is a Truth Value and How Many Are There?". *Studia Logica: An International Journal for Symbolic Logic*, Jul. 2009, Vol. 92, No. 2, Truth Values. Part II (Jul., 2009), pp. 183-201

Cotnoir, Aaron. "Nagarjuna's Logic," in Graham Priest; Koji Tanaka; Y Deguchi; and Jay Garfield (eds), *The Moon Points Back* (Oxford: Oxford University Press), 176-188. 2015.

D'Agostino, Marcello and Solares-Rojas, Alejandro. "Towards Tractable Approximations to Many-Valued Logics: the Case of First Degree Entailment." 2022.

David, Marian, "The Correspondence Theory of Truth", *The Stanford Encyclopedia of Philosophy* (Winter 2020 Edition), Edward N. Zalta (ed.),
URL = <https://plato.stanford.edu/archives/win2020/entries/truth-correspondence/>.

Dembroff, Robyn. “What is Sexual Orientation?” *Philosophers’ Imprint* 16, no. 3 (January 2016)

Dunn, J. Michael. “Intuitive semantics for first-degree entailment and “coupled trees””, *Philosophical Studies* 29:149–168, 1976.

Erard, Michael. "The Life and Times of 'Colorless Green Ideas Sleep Furiously'". *Southwest Review* 95, no. 3 (2010): 418–425.

Field, Hartry. “Truth and the Unprovability of Consistency.” *Mind* 115, no. 459 (2006): 567–605.

Field, Hartry. “Tarski’s theory of truth.” In *The nature of truth: Classic and contemporary perspectives*, edited by Michael P. Lynch, 365-396. MIT Press, 2001.

Field, Hartry. “Deflationist views of meaning and content.” *Mind* 103 (411): 249-285. 1994.

Font, Josep. “Belnap's Four-Valued Logic and De Morgan Lattices.” *Logic Journal of the IGPL*. 5. 1-29. 1997.

Franchella, Miriam. ‘In the footsteps of Julius König's paradox’, *Historia Mathematica*, Volume 43, Issue 1, 65-86. 2016.

Frege, Gottlob. “Der Gedanke. Eine Logische Untersuchung”, translated as “Thoughts”. In *Collected Papers on Mathematics, Logic, and Philosophy*, edited by Brian McGuiness, 351-372. Oxford: Blackwell, 1984.

Garfield, Jay, *The Fundamental Principles of the Middle Way: Nāgārjuna’s Mūlamadhyamakakārikā*. Oxford University Press, New York. 1995

Garfield, Jay and Priest, Graham. “Mountains are just mountains.” In *Pointing at the Moon: Buddhism, Logic, Analytic Philosophy*, edited by D'Amato, Mario ; Garfield, Jay L. and Tillemans, Tom J. F. Oxford University Press, 71-82. 2009

Gerson, Lloyd, "Plotinus", *The Stanford Encyclopedia of Philosophy (Fall 2018 Edition)*, Edward N. Zalta (ed.),
URL = <https://plato.stanford.edu/archives/fall2018/entries/plotinus/>.

Goddard, Leonard and Routley, Richard. The Logic of Significance and Context, Vol. 1. Edinburgh, Scotland: Scottish Academic Press. 1973

Goddard, Cliff and Wierzbicka, Anna. *Words and Meanings: Lexical Semantics Across Domains, Languages, and Cultures*. OUP, 2013.

Hallden, S. “The logic of nonsense,” Ph.D. dissertation, Uppsala ´ University, 1949.

Hansen, Chad, "Daoism", *The Stanford Encyclopedia of Philosophy* (Spring 2020 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/spr2020/entries/daoism/>.

Haslanger, Sally. “Gender and Race: (What) Are They? (What) Do We Want Them to Be?”. *Noûs* 34, no. 1 (2000): 31-55.

Heck, Riki. “Truth and disquotation.” *Synthese* 142 (3):317-352. 2005.

Horwich, Paul. “A defense of minimalism.” In *The nature of truth: Classic and contemporary perspectives*, edited by Michael P. Lynch, 559-577. MIT Press, 2001.

Ibsen, Henrik. *Peer Gynt.* Translated by William and Charles Archer. Project Gutenberg, 2006. URL=<https://gutenberg.net.au/ebooks06/0608461.txt.>

Jayatilleke, K. N. “The Logic of Four Alternatives.” *Philosophy East and West*, vol. 17, no. 1/4, 69-83.1967.

Jonas, Silvia. *Ineffability and its Metaphysics: The Unspeakable in Art, Religion, and Philosophy*. Palgrave Macmillan New York. 2016

Kleene, Stephen Cole. "On Notation for Ordinal Numbers", *The Journal of Symbolic Logic*, Association for Symbolic Logic, **3** (4): 150–155. 1938.

Kleene, Stephen Cole. *Introduction to Metamathematics*. Princeton, NJ, USA: North Holland. 1952.

Kripke, Saul. “Outline of a theory of truth.” *Journal of Philosophy* 72 (19):690-716. 1975.

Łukasiewicz, Jan. ‘Philosophical remarks on many-valued systems of propositional logic.’ Translated by H. Weber. In *Polish logic 1920-1939*, edited by Storrs McCall. The Clarendon Press, Oxford, 40–65. 1967.

Lynch, Michael P. “Realism and the Correspondence Theory: Introduction.” In *The nature of truth: Classic and contemporary perspectives*, edited by Michael P. Lynch, 9-15. MIT Press, 2001.

Makinson, David Clement. *Topics in Modern Logic*. London: Methuen; Distributed by Harper & Row Publishers, Inc., Barnes and Noble Import Division. 1973.

Mares, Edwin, "Relevance Logic", *The Stanford Encyclopedia of Philosophy* (Winter 2020 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/win2020/entries/logic-relevance/>.

Maudlin, Tim. *Truth and Paradox: Solving the Riddles*. OUP, 2004.

Mendelson, Elliott. *Introduction to Mathematical Logic (6th edition)*. Routledge, 2015.

Millar, Robert McColl and Larry Trask. *Trask's Historical Linguistics (3rd edition)*. Routledge, 2015.

Moore, G. E. "Moore's Paradox". In *G. E. Moore: Selected Writings*, edited by Thomas Baldwin, 207-212. London: Routledge, 1993.

Næss, Arne. *"Truth" as Conceived by Those Who Are Not Professional Philosophers*. Skrifter Utgitt av Det Norske Videnskaps-Akademi I Oslo Il. Hist.-Filos. Klass No. 4. Oslo: I Komisjon Hos Jacob Dybwad. 1938.

OED online. "chair, v.". OUP, March 2022. Accessed April 21, 2022. https://www-oed-com.ezproxy.uio.no/view/Entry/30217?rskey=9lm5Yt&result=3&isAdvanced=false.

Omori, Hitoshi. "Hallden's Logic of Nonsense and Its Expansions in View of Logics of Formal Inconsistency," *2016 27th International Workshop on Database and Expert Systems Applications (DEXA)*, 129-133. 2016.

Omori, Hitoshi and Wansing, Heinrich. "40 years of FDE: An Introductory Overview." *Studia Logica* 105 (6):1021-1049. 2017

Pietz, Andreas and Rivieccio, Umberto. "Nothing but the Truth." *Journal of Philosophical Logic* 42 (1):125-135. 2013.

Plotinus. *The Divine Mind, Being the Treatises of the Fifth Ennead*, trans. Stephen Mackenna and B. S. Page. Boston, MA: Charles T. Branford Company. 1918.

Priest, Graham. "The Logic of Paradox", *Journal of Philosophical Logic*, 8(1): 219–241. 1979.

Priest, Graham. *An Introduction to Non-Classical Logic: From If to Is*. Cambridge University Press, 2008.

Priest, Graham. “Many-valued modal logics: a simple approach.” *The Review of Symbolic Logic* 1, no. 2: 190–203. 2008.

Priest, Graham. “The logic of the catuskoti”. *Comparative Philosophy* 1 (2):24-54. 2010.

Priest, Graham. “Plurivalent logics”. *Australasian Journal of Logic* (11), Article no. 1: 2-13. 2014.

Priest, Graham. ‘None of the Above: The Catuṣkoṭi in Indian Buddhist Logic’, in *New Directions in Paraconsistent Logic*, 517-527. 2015.

Priest, Graham. '*Quintum Non-Datur*', *The Fifth Corner of Four: An Essay on Buddhist Metaphysics and the Catuskoti*. Oxford Academic; online edn, 2018.

Priest, Graham. “Natural Deduction Systems for Logics in the *FDE* Family”. In: Omori, H., Wansing, H. (eds) *New Essays on Belnap-Dunn Logic*. Synthese Library, vol 418. Springer, Cham. 2019.

Radhakrishnan, S., Moore, C. (eds.): *A Source Book in Indian Philosophy*. Princeton University Press, Princeton. 1957

Rescher, Nicholas. *Many-Valued Logic*. New York: Mcgraw-Hill. 1969.

Riga, Peter J. “On Truth: A Catholic Perspective.” *Philosophy East and West* 20, no. 4 (1970): 369–76.

Routley, Routley, and Val Routley. “The Semantics of First Degree Entailment.” *Noûs* 6, no. 4: 335–59. 1972.

Scharp, Kevin. *Replacing Truth*. Oxford University Press, 2013.

Shaw-Kwei, Moh. “Logical Paradoxes for Many-Valued Systems.” *The Journal of Symbolic Logic*, vol. 19, no. 1, 37-40. 1954.

Sheffer, Henry M. “Ineffable Philosophies.*” The Journal of Philosophy, Psychology and Scientific Methods* Vol. 6, No. 5: pp. 123–129. 1909.

Shramko, Yaroslav and Heinrich Wansing, "Truth Values", *The Stanford Encyclopedia of Philosophy* (Winter 2021 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/win2021/entries/truth-values/>.

Simmons, Keith. *Universality and the liar*. Cambridge University Press, 1993.

Soames, Scott. "What is a theory of truth?". In *The nature of truth: Classic and contemporary perspectives*, edited by Michael P. Lynch, 397-418. MIT Press, 2001.

Stefon, Matt. "dao". *Encyclopedia Britannica*, 27 Sep. 2022, https://www.britannica.com/topic/dao. Accessed 13 December 2022.

Szmuc, Damian and Omori, Hitoshi. "A Note on Goddard and Routley's Significance Logic." *Australasian Journal of Logic* 15 (2):431-448. 2018.

Talmy, Leonard. "Semantics, universality of." In *The Cambridge Encyclopedia of the Language Sciences*, edited by Patrick Colm Hogan. Cambridge University Press, 2011.

Tarski, Alfred. "The concept of truth in formalized languages". In *Logic, Semantics, Metamathematics: Papers From 1923 to 1938*, 152-278. New York, USA: Hackett, 1956.

Tarski, Alfred. "The semantic conception of truth". In *The Nature of Truth Classic and Contemporary Perspectives*, edited by Michael P. Lynch, 331-363. MIT Press, 2001.

Ulatowski, Joseph. "Ordinary truth in Tarski and Næss." In *Uncovering facts and values*, edited by Adrian Kuzniar and Joanna Odrowąż-Sypniewska, 67-90. Brill, 2016.

Wittgenstein, Ludwig. *Tractatus Logico-Philosophicus. Logisch-philosophische Abhandlung*. Frankfurt a.M.: Suhrkamp. 1922

Young, James O., "The Coherence Theory of Truth", *The Stanford Encyclopedia of Philosophy* (Fall 2018 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/fall2018/entries/truth-coherence/>.

Zhang, Longxi. *The Tao and the Logos: Literary Hermeneutics, East and West*, Duke University Press, 1992.