Pleitz · Logic, Language, and the Liar Paradox

Martin Pleitz

Logic, Language, and the Liar Paradox

mentis
MÜNSTER

Einbandabbildung: Sandzeichnung mit kreisförmigem Pfeil (Foto: Martin Pleitz)

Bibliografische Information der Deutschen Nationalbibliothek

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

D6: Philosophische Dissertation, Universität Münster

Gedruckt auf umweltfreundlichem, chlorfrei gebleichtem
und alterungsbeständigem Papier ♾ ISO 9706

Eisenbahnstraße 11, 48143 Münster, Germany
www.mentis.de

Printed in Germany
Einbandgestaltung: Anna Braungart, Tübingen
Wissenschaftlicher Satz: satz&sonders GmbH, Münster (www.satzundsonders.de)
Druck: AZ Druck und Datentechnik GmbH, Kempten
ISBN 978-3-95743-034-2 (Print)
ISBN 978-3-95743-849-2 (E-Book)

In memory of Rosemarie Rheinwald.

Preface

In this book I discuss the Liar paradox and propose a novel solution to it. The Liar paradox arises in the following way: We consider a sentence which says of itself that it is false, or a sentence which says of itself that it is not true. If such self-referential sentences exist (and examples such as 'This sentence is false' and 'The sentence on the whiteboard in room 101 is not true' which is written on said whiteboard certainly suggest this), then our logic and standard notion of truth force a contradiction upon us: The Liar sentence is true and false! Or even: The Liar sentence is true and not true! What has gone wrong? Must we revise our notion of truth and our logic? Or can we dispel the common conviction that there are such self-referential sentences? Most people today take the first way because they think that the second way is blocked. But this is the route I have taken.

Here is a brief report of where the road has taken me: By comparing the Liar reasoning in formal and informal logic (in part I), I have shown that any adequate discussion of the Liar paradox needs to construe the expressions involved as meaningful, even in cases where they belong to a language of logic. For this reason, I have turned to the theory of meaning (in part II). Based on a detailed overview of the semantics of singular terms, I have given a comprehensive list of the (purported) means for constructing self-referential sentences. I have also shown that even in the mediated sense in which the method of Gödelization does allow to construct sentences that attribute certain features (like unprovability) to themselves, it cannot be used to construct Liar sentences – despite a common misconception. After these extensive preparations, I have developed my own account of the metaphysics and semantics of self-referential languages (in part III). According to it, meaningful expressions are distinct from their syntactic bases and from their physical bases, and they exist only relative to contexts. By recourse to several semantico-metaphysical arguments I was able to show that in this dynamic setting, an object can be referred to only *after* it has started to exist. Hence self-reference of the kind that leads to the Liar paradox cannot occur – despite our initial impression. As this blocks the paradoxical reasoning at an early point, we are freed from the need to revise our notion of truth and our logic in a major way. And as this solution is contextualist, it evades the expressibility problems of other proposals.

My main aim in writing about the Liar paradox was of course to convince the experts working in the field – philosophers and logicians who are worried about the paradox. But due to the particular character of my proposal, the investigation also needed to involve three more general questions. The first two of these belong to philosophical logic and the philosophy of language, respectively:

What is the character of a language of logic?
How does semantics interact with the metaphysics of expressions?

The third question is probably more surprising in a study about the Liar paradox, because it belongs to the philosophy of time:

How can tensed reality be represented from within?

Only on the basis of answers to these more general questions was I able to give my particular answer to the more specific question:

How can the Liar paradox be solved?

In other words, I have combined insights from philosophical logic, semantics, the metaphysics of expressions, and the tensed theory of time when I developed my extended argument for the conclusion that the circular reference needed in the Liar paradox cannot occur. Therefore this book should be of some interest also to philosophers who do not themselves work on the Liar paradox.

For these reasons, the present study has grown into a long text. As the first chapter gives an introduction to the problem and explains the plan of the book, this preface need not contain anything more by way of preview. But let me say something about how you can read the book. Obviously, an ideal reader would read all of it, and then once more! But I am of course aware that few people will be able and willing to allot that much time to working through the results of my work. Therefore I have written part III, The Metaphysics of Expressions and the Liar Paradox, in a way that should enable you to assess the argument for my proposed solution on the basis of reading only this last part of the study, following back the references to parts I and II as needed. What is more, I would not discourage an impatient reader who is already an expert on these matters from jumping right to chapter 13, Subsequentism Solves the Liar Paradox: Reading only this last chapter would probably allow you to understand the basics of my proposal – although not to criticize it fairly.

Acknowledgements

First, I would like to express my gratitude to some of the people who are important to me for reasons not directly connected to the project of writing this book. I am grateful to my parents for their love and their trust in my abilities, and I thank Simon, Dario, Jennie, Roman, Tobi, Pete, Deni, Arda, Fynn, and Daniel for making my life more beautiful.

Next, I would like to thank members of the Department of Philosophy at the University of Münster for their support, especially during a difficult phase after the untimely death of Rosemarie Rheinwald, who was my Ph.D. supervisor and employer then. First and foremost I want to mention Niko Strobach and Sibille Mischer, and in this regard I would also like to thank Peter Rohs, Oliver Scholz, Reinold Schmücker, and Kurt Bayertz, as well as Ansgar Seide, Eva Jung, Dirk Franken, and Christa Runtenberg. I am grateful to the University of Münster for awarding its Dissertation Prize to the Ph.D. thesis that this book is based on.

Many people have helped me write this book – by discussing the topic with me, by reading and commenting on preliminary drafts of the text, by encouraging me, and by assisting me with other tasks during the more intense phases of writing. I am grateful to all of them! Most important were discussions with Johannes Korbmacher at the outset of the writing process, the steady support by Niko Strobach throughout, and discussions with Tobias Martin towards the end. Simon Dickel helped me greatly during some crucial months in the winter of 2011/12. In these regards I would also like to thank Mathieu Beirlaen, Andreas Bruns, Chris Scambler, Daniel Boyd, Kit Fine, Toff Fromme, Severi Hämäri, Ulf Hlobil, Andreas Hüttemann, Katrin Huxel, Roman Klauser, Vojtěch Kolman, Nikola Kompa, Ernst-Wilhelm Krekeler, Roland Mümken, Christian Nimtz, Philipp Offermann, Vincent Peluce, Graham Priest, Lena Raezke, Stephen Read, Dave Ripley, Peter Rohs, Raja Rosenhagen, Stefan Roski, Liam Ryan, Sebastian Schmoranzer, Lionel Shapiro, Stewart Shapiro, Ralf Schindler, Ansgar Seide, Hartley Slater, Linda Supik, Vítězslav Švejdar, Allard Tamminga, Alex Thinius, Heinrich Wansing, Chrischi Wolf, Maiko Yamamori, and Elia Zardini.

Last but not least, I greatly appreciate my philosophy teachers. A long time ago, there were Rainer Schmidt and Marcus Willaschek. I would like to thank Rainer for getting me started and Marcus for encouraging me to go on. Today, I have the great fortune to find myself staying for a while with Graham Priest and Kit Fine in New York. I would like to thank Graham for an important conversation in Berlin and Kit for an important conversation in Oxford, and both of them for their generous support in many matters. The teacher, however, who the project of writing this book is most closely connected to is Rosemarie Rheinwald, who I continue to admire for her clarity, honesty, and kindness. When she fell ill, I re-read her book on the semantic paradoxes, but she died before I could tell her my ideas about it. I dedicate this book to her memory.

Table of Contents

PART I LOGIC AND THE LIAR PARADOX

Raymond Smullyan:
"This joke is not funny!"

Chapter 1 The Liar and Its Kind

Rosemarie Rheinwald:

"Dabei hängt dem Ergebnis häufig etwas Unerwartetes, Überraschendes an. Das Gefühl, das eine Paradoxie auslöst, ist ein Gefühl der Verwirrung."[1]

The Liar paradox poses a serious challenge for logic, and it prompts us to think deeply about language – about the meaning of linguistic expressions as well as about their metaphysical nature. In fact, paying close attention to how the metaphysical nature of expressions interacts with their meaning will allow us to solve the paradox. Or so I will argue in the present study about *Logic, Language, and the Liar Paradox*.

The study is divided into three parts, corresponding to the three philosophical subdisciplines of logic, semantics (which is concerned with the meaning of expressions), and metaphysics – in particular, the metaphysics of expressions (which is concerned with the metaphysical nature of expressions). The aim of the first part is to look at the Liar paradox from the viewpoint of logic, in order to assess how severe a problem it is, and of what kind. The present chapter introduces the Liar paradox and uses it as an occasion to clarify the notion of a paradox. The following chapters 2 and 3 will present the reasoning of the Liar paradox in detail from the angle first of informal logic and then of formal logic. In the second part of the study we will look at the Liar paradox from the viewpoint of semantics, and in the third part we will look at it from the viewpoint of the metaphysics of expressions, and solve it.

In this first chapter we will, after a look at the concept of a paradox in general (in section 1.1), present the Liar paradox (in section 1.2) and some of its close relatives (in sections 1.3 and 1.4). Then we will turn to the history of the Liar paradox: After a brief remark about its pre-modern phase (in section 1.6), we will concern ourselves with set theory and Cantor's diagonal argument (in section 1.7), which constitutes the starting point of the modern history of the Liar paradox, and led to interest in a larger group of paradoxes (in section 1.8). We will conclude the chapter with a quick review of some typical opinions about how problematic the Liar paradox is, which will raise the question whether it looks daunting only to those philosophers who prefer to work in a formal way (in sections 1.9 and 1.10).

---

[1] "And often, the result is unexpected and surprising. The feeling that is evoked by a paradox is a feeling of confusion." (Rheinwald 1988, 9; my translation)
The motto for part I is taken from Smullyan 2006, 152.

1.1 The concept of a paradox

Before turning to the Liar paradox, we should characterize the concept of a paradox in general. Willard Van Orman Quine asks: "May we say [...] that a paradox is just any conclusion that at first sounds absurd but has an argument to sustain it?"[2] And with Quine, we can give an affirmative answer. Loosely speaking, any surprising and counterintuitive result of a given piece of seemingly sound reasoning can be called a 'paradox'.[3] We speak, for instance, of *the paradox of Achilles* (which is about the runner Achilles who cannot overtake the turtle although he is faster)[4] and, in connection with the theory of special relativity, we speak of *the twin paradox* (which is about two people who are of markedly different age although they were born on the same day). Clearly, either the paradoxical reasoning must be rejected despite its seeming soundness (as in the case of Achilles) or the paradoxical result must be accepted despite its initial counterintuitivity (as in the case of the twins). That is, a paradox is a problem that calls for a solution.

Now, in order to speak less loosely, the notions of *seemingly sound reasoning* and of *a surprising and counterintuitive result* should be made more precise. This is standardly done in logical terms, by representing a *piece of reasoning* as a collection of premises which form an argument, the conclusion of which stands in for the *result* of the reasoning. On this basis we can define the notions that we will work with:

(Def. Paradox) A *paradox* is an argument that appears to be valid from premises that appear to be true to a conclusion that appears to be unacceptable.[5]

(Def. Solution) A *solution* to a given paradox is an explanation[6] that dispels some of the appearances that are mentioned in (Def. Paradox) as illusory.

(Def. Antinomy) An *antinomy* is a paradox with a contradictory conclusion.

Some remarks on these definitions and the reasons for adopting them.

With regard to the definition of a paradox (and consequently that of an antinomy), there is some terminological leeway.[7] Often, it is not the argument as a whole but only its conclusion that is called a 'paradox' (or an 'antinomy').[8] But an unacceptable (or a contradictory) statement that stands alone is just that, unaccept-

[2] Quine 1976a, 1.

[3] Cf. Clark 2007, 151–154 and Sainsbury 1988, 1ff.

[4] Cf. Salmon 1970 and Rheinwald 2012[1993].

[5] This is just a slight terminological variation on a definition given by Greg Restall: "A *paradox* is a seemingly valid argument from seemingly true premises to a conclusion that is seemingly unacceptable" (Restall 1993, 281).

[6] More will be said about explanatoriness in sections 7.2 and 7.3.

[7] I would like to thank Severi Hämäri for alerting me to this point.

[8] Quine, Sainsbury, and Rheinwald by a 'paradox' mean the *conclusion* of a paradoxical argument. Cf. Quine 1976a, 1; Sainsbury 1988, 1; and Rheinwald 1988, 9.

able (or contradictory). To constitute a paradox, it must be backed by an argument. Hence, we should prefer the usage according to which a paradox is an argument, of course with a paradoxical conclusion.[9] Further reasons for preferring this usage will emerge later.[10]

The boundary between antinomies and paradoxes that are not antinomies is less clear – or rather, less *stable* – than might be expected. To be sure, the paradigmatic formulations of the Liar paradox and Russell's paradox are antinomies.[11] And equally clearly, the standard version of the paradox of Achilles and the twin paradox are not antinomies, because the paradox of Achilles is standardly construed as an argument with a non-contradictory conclusion that is false (that Achilles cannot overtake the turtle) and the twin paradox is standardly construed as an argument with a non-contradictory conclusion that *seems* to be false (that twins can be of markedly different age). But it is possible in both cases to generate an antinomic variant by adding the negation of the usual conclusion as a further premise. Zeno, for instance, had he argued differently, might have concluded that Achilles *can and cannot* overtake the turtle. And even the Liar paradox has a non-antinomic variant which concludes with the negation of one of the usual premises.[12] In fact, every antinomy can be transformed into a paradox that is not an antinomy and vice versa, in such a way that the transformation preserves the validity of the argument.[13] The reason for the instability of the status of being an antinomy is that the *individuation* of paradoxes is rather coarse-grained, because the same paradox can come in the form of different arguments.[14] To belong to the same paradox, these variants of course have to form a close-knit[15] family. (The possibility of reformulating a paradox by moving on to an argument that is in a sense equivalent constitutes another reason to let arguments and not their conclusions be the range of application for the notion of a paradox.)

It seems to be a common impression that antinomies are problems of a more severe nature than non-antinomic paradoxes, presumably because *being contradictory* is seen as a worse predicament for a claim or sentence to be in than simply *being false*. But *unacceptable* is *unacceptable*: It is by no means the case that a theory (say) with a false consequence is somehow more acceptable than an otherwise similar

---

[9] Mackie and Restall by a 'paradox' mean an *argument* with a paradoxical conclusion. Cf. Mackie 1973, 238; and Restall 1993, 281.

[10] Cf. sections 1.4 and 1.5.

[11] Cf. sections 1.2, 1.6, and 1.8.

[12] To anticipate, cf. steps (1) through (10) of Informal Variant 5 of the basic Liar reasoning in section 2.4.

[13] In any not too non-standard logic, a valid argument from the premise *that p* and a collection Γ of other premises to a contradiction can be recombined to a valid argument from the other premises Γ to the conclusion *that it is not the case that p*. And a valid argument from a collection Γ of premises to the conclusion *that p* can be recombined to a valid argument from those premises and the additional premise *that it is not the case that p* to a contradictory conclusion.

[14] To anticipate, we will look at several distinct variants of the basic Liar reasoning in section 2.4.

[15] We should not try to be very precise on this point; *paradoxology* should not be carried too far. Paradoxes, after all, are *mistakes* and intimately tied to what must be false appearances, so that they make for an evasive subject.

theory with a contradictory consequence – in both cases there is a problem that *must* be solved.[16]

This brings us to a question which will later turn out to have an interesting answer: *What* is it that makes a conclusion of a paradox appear to be unacceptable? We have seen two clear cases, the appearance of falsity and the appearance of contradictoriness. But are these all the options? Later on we will see that they are not, because there are arguments with a true conclusion which strike us as paradoxical.[17] (Here it should be enough to remind us of the *paradoxes of material implication*: Because of the classical understanding of implication, every claim – and hence every *true* claim – is implied by any contradictory claim. If we want to classify each one of the arguments of this form as paradoxical, we already have found many paradoxes with a true conclusion.[18])

*

Now let us turn to the relation between a paradox and its solution. Mark Sainsbury is of course right when he observes, with regard to a definition of a paradox similar to the above (Def. Paradox):

> "Appearances have to deceive, since the acceptable cannot lead by acceptable steps to the unacceptable."[19]

This means nothing else than that a paradox *must be solved*, and it motivates a definition of a paradox's solution like the above (Def. Solution), which makes explicit its close connection to the definition of a paradox, insofar as it requires of a solution that it removes "some of the appearances that are mentioned in (Def. Paradox)".

This close connection also is a reason to adopt the above (Def. Antinomy), which is a slight deviation from the following classical definition:

> A *classical antinomy* is an argument that appears to be valid from premises that appear to be true to a contradictory conclusion.[20]

The difference to the notion of an antinomy in our sense becomes evident when we note that (Def. Antinomy) and (Def. Paradox) together entail the following principle:

> An antinomy is an argument that appears to be valid from premises that appear to be true to a contradictory conclusion that appears to be unacceptable.

As every contradictory conclusion is a contradictory conclusion that appears to be unacceptable[21] (and, of course, vice versa), an argument is a classical antinomy just in case it is an antinomy in the above sense. But this extensional equivalence

---

[16] And of course, a difference in severity would not go well with the possibility just noted of moving to and fro between the status of being an antinomy and that of being a non-antinomic paradox.

[17] Cf. sections 1.4 and 1.5.

[18] Cf. Sorensen 2005, 105f.

[19] Sainsbury 1988, 1.

[20] Cf. Haack 1978, 138; Rheinwald 1988, 9.

[21] Presuming that even dialetheists, i. e., logicians who accept some contradictions (Priest 2006b, 1; cf. Berto/Priest 2010), will concede that every contradiction at least *appears* to be unacceptable.

notwithstanding, there is an important difference on the conceptual level, because the classical notion of an antinomy leads naturally to a notion of an antinomy's *solution* that is not up to date. Given the classical definition of an antinomy, it is only natural to parallel the above definition of a solution to a paradox and characterize the corresponding notion of a solution to an antinomy by requiring that it dispel some of the appearances mentioned in the classical definition as illusory. Thus, a solution to an antinomy in our sense will consist either in showing the argument to be not valid, or in arguing against the truth of some of its premises, or in giving reasons for its contradictory conclusion to be acceptable, after all. By contrast, the solution to a classical antinomy will consist either in showing the argument to be not valid, or in arguing against the truth of some of its premises – but it is no option to give reasons for the acceptability of the contradictory conclusion. Of course, most people (including myself) will have no problem with this, deeming any contradiction to be unacceptable, period. But by now there is an important group of proposed solutions to the Liar paradox (and related paradoxes) which are based on just that – arguing for the acceptability of some contradictions.[22] So we should not exclude these proposals from the range of acceptable candidates for a solution by the very definitions of a paradox and of an antinomy.

*

Let us conclude the present section with some clarifactory remarks about the *appearances* that play a central role in the definition of a paradox (and therefore also in the connected definitions of a paradox's solution and of an antinomy). Firstly, the appearances should be *shared as a package* in the sense that it is (in general) *to the same people* that the premises seem true, the argument seems valid and the conclusion seems unacceptable. Otherwise, any conflict of intuitions among some philosophers would have to be characterized as a paradox.[23] Secondly, the appearances should be *intersubjective* in the weak sense of being *widely shared*.[24] If, e. g., the only one who finds some conclusion unacceptable is me, then the corresponding argument is certainly no paradox. Thirdly, there is an element of *historicity* because the status of being a paradox is changeable. Some paradoxes have, after all, been solved! When enough people change their minds, some of the premises might lose their appearance of truth, the reasoning might lose its appearance of validity, or the conclusion might lose its appearance of unacceptability. It has often been argued, e. g., that the paradox of Achilles today is a paradox in no more than name.[25]

There seems to be some tension between the aspects of (weak) intersubjectivity and of historicity because historicity precludes *diachronic* intersubjectivity. But

[22] We will have a brief look at the approach to the Liar paradox of Graham Priest's and other paraconsistent logicians in section 7.5.

[23] I would like to thank Roland Mümken for a conversation which alerted me to this point.

[24] "Paradoxes are fun. In most cases, they are easy to state and immediately provoke one into trying to 'solve' them." (Sainsbury 1988, 1)

[25] The matter is contentious. While a majority think that the Achilles paradox is solved, there remains a minority of dissenters. See the anthology Salmon 1970 and compare the different assessments in Rheinwald 1988, 11 and Rheinwald 2012[1993], 27.

the tension is only apparent. A paradox is usually connected (via the *seemingly sound reasoning* behind the paradoxical conclusion) to some *theory*. And, as of now, every theory has been generally accepted only during some phase in history; but during that phase it *did* have an intersubjective character (because it was *generally* accepted).

What we also find here is a kind of *democratic* element (in a loose sense of the word): The status of being an (unsolved) paradox is connected to what the majority of people (or more likely, a majority of *experts*) think at a time. E. g., the main reason for the contemporary near-consensus that the paradox of Achilles has been solved arguably is the impact that mathematical progress had on how the majority of people (that is, the majority of the people who care) today think about infinity.[26]

1.2 The Liar paradox

The Liar paradox in its typical modern variants that are discussed in philosophy and logic involves a *Liar sentence*, i. e., a sentence which says of itself that it is false or which says of itself that it is not true.[27] For reasons that will become clear later,[28] a sentence that says of itself that it is false is called a *simple* Liar sentence and a sentence that says of itself that it is not true is called a *strengthened* Liar sentence.

Unsurprisingly, the Liar paradox is indeed a paradox: When we, with respect to some Liar sentence, ask whether it is true, straightforward reasoning from plausible principles about truth and falsity will lead to a contradiction. Here is one typical way of arguing.

---

[26] From a philosophy of science point of view, paradoxes have often been praised for the incentive they provide for scientific progress, as foolproof signs of a serious crisis (cf. Bromand 2001, 11f.). I do not want to dispute this as an historical observation about the role paradoxes play in modern sciences. But, pace some philosophers and historians of science, this feature surely is not essential to the *concept* of a paradox. To anticipate section 1.6, in medieval times paradoxes like the Liar paradox were discussed extensively, but without a sense of serious crisis.

[27] In contrast to all of its contemporary academic variants, many popular (e. g., Clark 2007, 112) and most ancient (cf. Bocheński 1956, 151f. and Brendel 1992, 21f.) variants of the Liar paradox are formulated not in terms of falsity or untruth, but of *lying*. They concern a person who calls himself or herself a liar or a sentence that says of itself that it is a lie, e. g., 'What I am saying now is a lie'. Now roughly, to lie is to say something false (or untrue) with an intention to deceive (cf. Künne 2013, 24ff. for an extensive discussion; cf. Shibles 1988 for a diverging opinion). The element of deception, however, plays no role in any of the usual variants of the Liar reasoning. Thus the term 'Liar paradox' strictly speaking is a misnomer. For this reason, Wolfgang Künne proposes to change the name of the paradox to "the antinomy of falsity" or "f-antinomy" (Künne 2013, 39; my translation). But such attempts at language reform seem hopeless to me. Talk about "the Liar" is well-entrenched in the logic community. And this may be more than a slight terminological carelessness, for as long as the paradox appears so difficult to solve, the connotation of malign intentions feels apt, and is also in keeping with talk about the "*Revenge*" of the Liar (cf. section 2.6).

[28] Cf. section 2.5.

(Step 1) Let us say that we are given a simple Liar sentence,
i. e., a sentence which says of itself that it is false.
Now we ask: Is it true? Is it false?

(Step 2) We reason: Suppose that our simple Liar sentence is true.
Surely, if a sentence is true then what it means is the case.
Hence our simple Liar sentence must be false.
It follows that if our simple Liar sentence is true then it is false.

(Step 3) We reason: Suppose that our simple Liar sentence is false.
Surely, if what a sentence means is the case then the sentence is true.
Hence our simple Liar sentence must be true.
It follows that if our simple Liar sentence is false then it is true.

(Step 4) In sum, our simple Liar sentence is true if and only if it is false.

This rather intuitive way of starting the Liar reasoning might suggest that our Liar sentence moves back and forth between truth and falsity, being true at a first stage, hence false at a second stage, hence true at a third stage, hence false at a fourth stage, and so on *ad infinitum*. And that would be odd, especially as we could not appeal to any element *external* to the Liar sentence to explain these changes.[29] But the situation is even worse. On the usual logical understanding of 'if … then …'-constructions, there is no leeway for any temporal (or otherwise contextual) variation. If we give a temporal reading to a sentence of the form 'if p then q' at all, then it has to be 'if p, then *simultaneously* q'. Similarly, we would have to understand the claim we just have argued for as meaning that our simple Liar sentence is true if and only if it *simultaneously* is false. So we are well advised to set aside again any temporal or contextual understanding and turn back to the Liar reasoning.

For we are not finished yet, because only given certain assumptions about how truth relates to falsity will the (simultaneous) truth and falsity of a sentence be contradictory.[30]

(Step 5) If we add the further plausible assumption
that every sentence is true *or* false, we can infer:[31]
Our simple Liar sentence is true *and* false.

(Step 6) But, surely, no sentence is both true and false. Hence, we can infer:
Our simple Liar sentence is true and false and it is *not* true and false.
And that is as contradictory as it gets.

We can see that the essential ingredients of the basic[32] Liar reasoning are principles about truth and falsity, rules of logic, and a fact about language. More specifically,

[29] We can appeal to an external aspect, the flow of time, in the case of a sentence like 'Now it is night', which does seem to switch its truth value two times a day (Hegel 1988, 71).

[30] Cf. also the remark after Informal Variant 1 in section 2.4.

[31] To anticipate, we use the rule of reasoning by cases; cf. section 2.1.

[32] The reasoning about a Liar sentence that leads to a contradiction is *basic* in so far as it will turn out that there also is an *extended* Liar reasoning, which unfolds only after a first attempt at solving the paradox has been made. Cf. sections 2.5 through 2.10.

we have appealed to the principle that a sentence is true if and only if what it means is the case in (Step 2) and (Step 3) of the above presentation[33] and to the principles that truth and falsity are exhaustive and exclusive in (Step 5) and (Step 6).[34] We have employed certain logical rules in (Step 2) through (Step 6). Both the principles and the rules are in accordance with classical logic (and many other logical systems).[35]

The fact about language that is essential to the Liar reasoning is presupposed already in (Step 1). It is that there are indeed self-referential sentences and, in particular, that there are Liar sentences. Prima facie, there are many candidates for singular terms that can be used to form self-referential sentences. Here are some of them in action, i. e., occurring in sentences that are plausible candidates for Liar sentences:[36]

(1) 'This sentence is false.'

(2) 'Larry is false', under the assumption that it is named 'Larry'.

(3) '(3) is false.'

(4) 'The fourth sentence that is mentioned in the list of five sentences in section 1.2 of *Logic, Language, and the Liar Paradox* by Martin Pleitz is false.'

(5) 'The only sentence written on the whiteboard in room 101 at t_0 is false', under the assumption that that sentence is the only sentence written on the whiteboard in room 101 at t_0.

Each one of the five sentences mentioned here seems to say of itself that it is false, i. e., each one seems to be a simple Liar sentence.

Examples like these and many others have convinced almost everyone currently working in the field of philosophical logic that at least some devices of sentential self-reference are unproblematic. Therefore the relevance of the Liar paradox is seen today by most people to lie in the fact that it calls into question principles about truth and falsity we commonly endorse, or logical rules we commonly employ, or both. For the time being I will refrain from challenging this near-consensus of contemporary logicians. But let me note at the outset that my own proposal for a solution to the Liar paradox will consist in an argument to the conclusion that – despite appearances – the required kind of self-reference is not possible, after all.[37]

Because of this aim, we will take a different way than many current presentations when discussing the other essential ingredients of the Liar paradox, our notion of truth and the logic of our reasoning. For those current paradox-solvers who are convinced that there is the kind of self-reference required for the Liar paradox, a discussion of truth and logic must be *revisionary*, i. e., it must concern how the principles about truth or the rules of logic should be *changed* to block the paradoxical

---

[33] Cf. section 2.2.

[34] Cf. section 2.3.

[35] Cf. sections 2.1 and 3.3.

[36] For better readability, I often omit the period when mentioning a single sentence.

[37] In chapter 5 we will investigate more thoroughly the different ways sentential self-reference seems to be achievable; in chapter 6 we will deal with the widespread conviction that the technique of Gödelization allows to form a formal Liar sentence; and in part III, we will argue against the possibility of sentential self-reference of the sort needed to construct Liar sentences.

arguments. These ways of dealing with paradox have developed into a rich field of inquiry.[38] In contrast, we do not need to concern ourselves overmuch with logical revision in the present study about the Liar paradox. There might of course be good reasons to explicate[39] our notion of truth in a non-standard way or to explicate our reasoning in a way that deviates from classical logic. But if I am right, the Liar paradox and its relatives are not among these reasons. On our way to show that, we will of course look at current explications of truth and logic – but not with a view of revising them, but rather to ask what they are, and how they relate to each other and to our understanding of language. As these questions are interesting in their own right, our aim in the present study really is twofold: We want to use the Liar paradox as an occasion to think about the nature of logic and the nature of language, and we want to solve it.

1.3 Liar cycles and Yablo's paradox

In the last section we had a first glance at the variety of options for singular terms which seem to allow the construction of self-referential sentences. Now, we will turn to systems of more than one sentence which achieve Liar-like results: the *finite loops* used to form Liar cycles and the *infinite sequences* which are the root of Yablo's paradox.

A Liar cycle is a construction of two or more, but finitely many, sentences which refer to each other in a way that leads to a situation as paradoxical as that engendered by a Liar sentence. An example medieval logicians were fond of concerns Plato and Socrates, who speak simultaneously:

Plato:	Socrates:
'What Socrates is saying right now is true.'	'What Plato is saying right now is false.'

There are Liar cycles of any number of sentences, as the following example of four sentences illustrates:

(1) '(2) is true.'

(2) '(3) is true.'

(3) '(4) is true.'

(4) '(1) is false.'

---

[38] Cf. section 7.5 for a concise overview of the spectrum of approaches to the Liar paradox, including approaches of logical revision.

[39] For the notion of *explication*, cf. Carnap 1956[1947], 7f. Carnap writes: "The task of making more exact a vague or not quite exact concept used in everyday life or in an earlier stage of scientific or logical development, or rather of replacing it by a newly constructed, more exact concept, belongs among the most important tasks of logical analysis and logical construction. We call this the task of explicating, or of giving an *explication* for the earlier concept; this earlier concept, or sometimes the term used for it, is called the *explicandum*; and the new concept, or its term, is called the *explicatum* of the old one." (Carnap 1956[1947], 7f.) In Quine's concise formulation, to *explicate* is to "devise a substitute, clear and couched in terms of our liking, that fills the functions" (Quine 1960, 259).

The purpose of examples like these – besides the fun they seem to stir in some people – is to show that sentential self-reference is not necessary for the Liar paradox (nor for paradoxes much like it). Some have even denied that the notion of the referential circularity involved in Liar cycles is *intelligible*.[40] But as we already have an understanding of what it is for a sentence to refer to an object,[41] it is easy to define what it is for a sentence to *refer** to an object, namely to be connected by an appropriate referential chain of sentences to the object. (In logical jargon, the relation of reference* is the *transitive closure* of the relation of reference.) And clearly, every sentence in a Liar cycle refers* to itself.

A more severe test for the claim that self-reference or at least some kind of circularity is at the root of Liar-like paradoxes is *Yablo's paradox*.[42] It concerns an infinite sequence of sentences, each one saying that all the following ones are false. One example would be the following *Yablo sequence*:

$S_1 =_{def}$ 'For every i greater than 1, S_i is false.'

$S_2 =_{def}$ 'For every i greater than 2, S_i is false.'

$S_3 =_{def}$ 'For every i greater than 3, S_i is false.'

...

Another Yablo sequence makes use only of indexicals:

'Every sentence in this sequence that is later than this one is false.'

'Every sentence in this sequence that is later than this one is false.'

'Every sentence in this sequence that is later than this one is false.'

...

Like Liar sentences, Yablo sequences produce paradox. For suppose that, for some number n, the n^{th} sentence is true. Then all following sentences must be false. But then the $n+1^{th}$ sentence would have to be true, and hence true and false. As this is a contradiction, the n^{th} sentence is false. And as this argument works for any number n, we can use it to show that *every* sentence in the Yablo sequence is false, and, in particular, that every sentence that follows the first is false. But then the first sentence must be true, hence true and false, which is a contradiction.

Although some people have tried to argue that there is some hidden circularity at work in Yablo's paradox, this is far from obvious – and the matter is highly contentious.[43]

As in the case of Liar sentences, it seems to be difficult to deny that there are Liar cycles and Yablo sequences. Therefore the paradoxes that they engender are

---

40 Cf. Leitgeb 2002.

41 To anticipate, we will say that a sentence *refers* to an object if and only if some singular term that is used in that sentence refers to that object. Cf. section 5.1, where we will say more about the relata of the relation of reference, primary and derivative.

42 Yablo 1993.

43 There is a small but lively debate about whether Yablo's paradox is circular in some sense; cf. Priest 1997, Sorensen 1998, Beall 2001, Leitgeb 2002, Cook 2004a & 2004b, Ketland 2005, and Yablo 2006. For a book-length study about Yablo's paradox, cf. Cook 2014.

important additions to the Liar paradox, which put further pressure on our common theories about truth and logic.

1.4 Church's paradox

Due to an early Christian's derogatory remark about the inhabitants of Crete being documented in the Bible, a certain variant of the Liar paradox has become popular according to which a Cretan utters the sentence 'All Cretans are liars'.[44] As that particular Cretan was identified as Epimenides of Knossos, this variant became known as "the paradox of Epimenides".[45]

To have a chance of being a starting point of a variant of the usual basic Liar reasoning, the utterance of Epimenides must be understood to mean that *every sentence* uttered by a Cretan is a lie – or rather, false. But that is not enough. For note that in the highly probable event that some other sentence uttered by a Cretan is true, the utterance of Epimenides, understood in this way, is not contradictory but just false. This has often been pointed out; for instance by Quine, who chided:

> "Actually, the paradox of Epimenides is untidy; there are loopholes. Perhaps some Cretans were liars, notably Epimenides, and others were not; perhaps Epimenides was a liar who occasionally told the truth; either way it turns out that the contradiction vanishes."[46]

Therefore, for the reasoning to be paradoxical in the usual way, the slightly slanderous[47] assumption is needed that all *other* sentences uttered by a Cretan (including all other sentences uttered by Epimenides) are false. As many people have treated the problem generated by the utterance of Epimenides simply as a variant of the Liar paradox, a principle of charity demands that we think of them as adding the slanderous assumption, albeit tacitly. Thus, *the Epimenides paradox* is the paradox that starts from the *two* assumptions, that, firstly, there is a sentence which means that all sentence of some sort are false, which is of that sort, and that, secondly, all other sentences of that sort are false.[48]

Treating the paradox of Epimenides as a contingent – and improbable – variant of the Liar paradox does not mean to belittle its paradoxicality. But Alonzo Church has observed that even without making the improbable extra assumption we are prone to fall into a deep problem when we ponder the utterance of Epimenides. Had not Arthur Prior and John Mackie taken it up later, Church's observation might well have gone unnoticed,[49] because he formulated it in only two sentences of a brief (but devastating) review of a text by Alexandre Koyré about the Liar paradox. There Church draws attention to a consequence of Epimenides making his utterance …

---

[44] Cf. Anderson 1978[1970], 1ff. – It is unknown *since when* the paradox of Epimenides is well-known; it was not discussed in the medieval insolubilia literature (Spade/Read 2009, section 1.2).

[45] Cf. Anderson 1978[1970], 2f., as well as the more recent and more detailed Künne 2013, 48–66.

[46] Quine 1976a, 6. – See also, e. g., Anderson 1978[1970], 3; Haack 1978, 136.

[47] We see that the anti-Cretan sentiment of the Biblical passage seeps through even to the logical level.

[48] For an explicit characterization, cf. Rheinwald 1988, 18f.

[49] Church 1946; Prior 1961; and Mackie 1973, 276ff.

> "[…] which, though not outright antinomy, might well be classed as paradox. Namely, without factual information about *other* statements by Cretans, it has been proved by pure logic (so it seems) that *some other* statement by a Cretan, not the famous statement of Epimenides, must once have been true."[50]

For suppose that some Cretan in fact utters 'Every sentence uttered by a Cretan is false'. Then we can turn the Liar reasoning of the paradox of Epimenides into a reductio argument for the claim that there is another, true, sentence uttered by a Cretan. We need only assume for reductio that all other Cretan utterances are false, follow the steps of the usual Liar reasoning[51] to infer a contradictory claim, and conclude that there must be some Cretan utterance that is true. Even in the very probable case that this conclusion is true, it leads to a further conclusion that is obviously unacceptable: The reasoning establishes a necessary connection between two distinct contingent states of affairs (for if there is the utterance of Epimenides, then it *must* be the case that there is some true Cretan utterance). In the words of Mackie, this "would violate Hume's principle";[52] and it should fly in the face even of those who are less skeptical of modality than Hume because it contradicts some plausible principles about linguistic utterances, especially regarding the high degree of compatibility between distinct possible utterances.

Let us call this unacceptable reductio argument *Church's paradox* – to distinguish it from the more Liar-like paradox of Epimenides. While the paradox of Epimenides rests on Epimenides's utterance *contingently* producing an antinomy, Church's paradox shows that the same utterance *necessarily* produces a paradox, albeit a non-antinomic one. As noted before,[53] its not being an antinomy need not make Church's paradox a problem less severe than the Liar paradox. (We will have another look at non-antinomic paradoxes in the next section.)

1.5 Curry's paradox

There is a relative of the Liar paradox that is problematic and interesting in its own right. It is *Curry's paradox* and involves what we can call a *Curry sentence*. One specific example for a Curry sentence is Curry Snow, which means that if Curry Snow is true, then snow is white. Surprisingly, reasoning about this sentence that appeals only to the same principle about truth that we have used in the Liar reasoning[54] and employs a few uncontentious logical rules allows to infer that snow is white. Even more surprisingly, we can in a similar way infer *any arbitrary claim*, be it true, false, or even contradictory, when we start from a Curry sentence with a consequent that expresses that claim.

---

[50] Church 1946, 131.

[51] E. g., (Step 1) through (Step 6) in section 1.2.

[52] Mackie 1973, 276.

[53] Cf. section 1.1.

[54] More specifically, in Curry's paradox we use the principle that says that a sentence is true if and only if what it says is the case, but we need not appeal to the principles that say that truth and falsity are exhaustive and exclusive.

Here is the reasoning involved in this inference for the case of the specific Curry sentence Curry Snow:[55]

(Step 1)	Let us say that we are given a specific Curry sentence, which says of itself that if it is true then snow is white. Let us call that sentence 'Curry Snow'.
(Step 2)	We reason: Suppose that Curry Snow is true. Surely, if a sentence is true then what it means is the case. Hence if Curry Snow is true, then snow is white. Now, as we have supposed that Curry Snow is true, it follows that snow is indeed white. So, we have used the assumption that Curry Snow is true to show that snow is white. In other words, we have shown: If Curry Snow is true, then snow is white.
(Step 3)	But that is exactly what Curry Snow means! And surely, if what a sentence means is the case, then that sentence is true. Hence Curry Snow must be true.
(Step 4)	In (Step 2) we have shown that if Curry Snow is true, then snow is white; in (Step 3) we have shown that Curry Snow is true. Taking these two claims together, we can infer: Snow is white.

We can see that, besides the existence of our Curry sentence, nothing is involved in the Curry reasoning but the principle that a sentence is true if and only if what it means is the case and a few logical rules. In particular, these are the rules which govern the logical behavior of classical conditionals and, under the classical understanding, of 'if … then …'-constructions of natural language.[56]

From the form of the Curry reasoning presented above it is obvious that any sentence can take the place of the sentence 'snow is white' as the consequent of a Curry sentence, e. g., the sentence 'snow is not white' or even the sentence 'snow is white and snow is not white'. Many presentations concern a claim that is more surprising than the one used here, more often than not a contradictory or false claim – "for effect", as Jc Beall says[57]. This can turn the Curry reasoning into,

[55] Anticipating some material from chapter 2, the reasoning can be given more concisely:

(1)	Curry Snow means that if Curry Snow is true, then snow is white.	meaning of the Curry sentence
(2)	Curry Snow is true.	assumption for conditional proof
(3)	If Curry Snow is true then snow is white.	(1), (2), naïve truth principle
(4)	Snow is white.	(2), (3), modus ponens
(5)	If Curry Snow is true then snow is white.	(2)–(4), conditional proof
(6)	Curry Snow is true.	(5), (1), naïve truth principle
(7)	Snow is white.	(5), (6), modus ponens

[56] To anticipate, these are the rules of modus ponens and conditional proof. Cf. sections 2.1 and 3.3.

[57] Beall 2009, section 2.3.

say, an easy proof that God exists.[58] As variants of Curry's paradox can be used to show any arbitrary claim, it has been seen by some as a problem even more severe than the Liar paradox. Anil Gupta and Nuel Belnap, e. g., state about the instances of the principle that a sentence is true if and only if what it says is the case that "they are inconsistent not only on a few isolated points; as Curry's paradox shows, they are *thoroughly* inconsistent".[59] Of course, given the logical rule of ex falso according to which any contradiction entails any arbitrary claim,[60] the Liar paradox in the end is no less inconsistent than Curry's paradox, because applying the ex falso rule to the contradiction that is inferred in the basic Liar reasoning also allows to infer any arbitrary claim and thus both paradoxes lead to the most severe form of inconsistency, *triviality*. But Curry's paradox allows to reach triviality *more directly* than the Liar paradox. And as the ex falso rule is not needed in the Curry reasoning, there are logical systems according to which the Liar paradox allows to infer only *some* contradictory claims whereas Curry's paradox still allows to infer *every* contradictory claim. Therefore Curry's paradox constitutes a serious extra problem for those approaches to the Liar paradox that are based on rejecting the ex falso rule.

*

But, evidence of seriousness notwithstanding, it is prima facie questionable whether Curry's paradox fits the notion of a paradox we work with.[61] We have defined a paradox as an argument that appears to be valid from premises that appear to be true to a conclusion that appears to be unacceptable; and an antinomy as a paradox with a contradictory conclusion.[62] The reason for doubting that Curry's paradox satisfies this condition is that in cases like the above example, there seems to be nothing unacceptable about the conclusion of the Curry reasoning. For, as we all know, snow is white! It is of course possible to regiment the meaning of the label 'Curry's paradox' such that it applies to a whole bundle of Curry arguments, starting from a collection of different Curry sentences, which together will allow to infer contradictory claims. Then the conclusion of Curry's paradox will indeed be a contradiction.[63] – But although this bundling strategy of construing Curry's paradox as an antinomy fits well with the indications of seriousness we saw before, it fails to explain the feeling of paradoxicality or absurdity we have even in the case of those single Curry arguments which end with something as uncontentious as the claim that snow is white. For is it not strange to infer from a meaning fact and a principle about truth, using only logical rules, a claim that is empirically falsifiable?

[58] Cf. Prior 1955, 177f.

[59] Gupta/Belnap 1993, 15.

[60] Cf. section 2.1.

[61] I would like to thank Elia Zardini for alerting me to this point. Cf. also Pleitz 2015a, 239f.

[62] Cf. section 1.1.

[63] Also, we can of course start with a Curry sentence with an unacceptable consequent, and if we want to make sure we land in a paradox in the strict sense, with a Curry sentence Curry Contradiction, which can be translated as 'If Curry Contradiction is true then (snow is white and snow is not white)'. But this reaction is unsatisfactory because it does not explain why *every* instance of Curry's paradox feels paradoxical.

And the feeling of paradoxicality should stay with us even when we regard those Curry arguments which allow to infer an instance of a logical law, starting for example from a Curry sentence that means that if it is true then (snow is white if and only if snow is white). The real problem about these Curry arguments does not lie in their conclusion (*as such*, or *in isolation*) being unacceptable, but in the fact that an acceptable conclusion is *reached* way too easily. As Roy Sorensen observes:

> "The paradox can be in *how* you prove something rather than in what you prove. This point causes indigestion for those who say that all paradoxes feature unacceptable conclusions. Their accounts are too narrow."[64]

Although I do agree with Sorensen that the conclusions of some paradoxical arguments are by themselves acceptable, I dare dissent from his diagnosis of indigestion. There is a way to achieve harmony between the intuitive paradoxicality of even those Curry arguments with harmless conclusions and the notion of a paradox – by being liberal concerning the meaning of the term 'unacceptable conclusion'. We need only say that the conclusion of an argument can be unacceptable in two ways: firstly, by the claim that is the conclusion being unacceptable *in itself* (usually because it is contradictory or obviously false), and secondly, by the claim that is the conclusion being unacceptable *as the conclusion* of the given argument. If a conclusion is unacceptable in this second way, then although the conclusion surely does follow, given the premises and the logical rules that are used, it nonetheless intuitively *does not seem to follow*. And this is what is paradoxical about the above Curry argument for the claim that snow is white, and about the Curry argument for the claim that if snow is white then snow is white.[65]

*

A brief remark on the history (and name) of Curry's paradox: In 1942, it was stated in a set theoretic formulation by Haskell Curry (hence the name) who used it to show that there can be paradox without negation.[66] In 1955, the paradox was restated in a meta-mathematical paper by Martin Hugo Löb, as a comment on his proof of what has become known as *Löb's theorem*.[67] (The reasoning in the proof of Löb's theorem bears a far-reaching similarity to the Curry reasoning, but Löb does not mention Curry in his text. One or the other of these two facts probably is the reason for Curry's paradox sometimes going under the name of 'Löb's paradox'[68].) Also in 1955, Curry's paradox made its way from mathematical logic to less formal parts

---

[64] Sorensen 2005, 106.

[65] Sorensen made the observation quoted above not with regard to Curry's paradox, but with regard to the paradoxes of material and of strict implication (Sorensen 2005, 105f.). The present strategy, according to which the unacceptability of the conclusion of a paradoxical argument need not be intrinsic to the claim that is the conclusion, obviously can be transferred from Curry's paradox to the paradoxes of implication. It explains, e. g., why the argument from the claim that one is prime and not prime to the claim that Socrates is famous can be classified as paradoxical despite the truth of its conclusion.

[66] Curry 1942.

[67] Löb 1955.

[68] E. g., Barwise/Etchemendy 1987, 23. Cf. Boolos/Burgess/Jeffrey 2002, 237f., where an instance of the Curry reasoning is used to motivate the proof of Löb's theorem, but Curry is not mentioned.

of philosophy, when Peter Geach and Arthur Prior explained it in (somewhat) less technical terms than Curry and Löb.[69] In recent years, Curry's paradox has received growing attention due to its importance in discussions about those approaches to the Liar paradox that depart from classical logic.[70]

Curry sentences, in contrast to Liar sentences, are a modern invention. It has been pointed out by Stephen Read, however, that there are forerunners among medieval paradoxes of validity.[71] Corresponding to the Curry sentence used above, we need only consider the following self-referential argument:

This argument is valid.

Hence, snow is white.

Simple reasoning about this argument will allow to infer that snow is white.

1.6 The history of the Liar paradox

Some problems have been of interest to philosophers during all of European and Western history, but the Liar paradox was discussed only in three phases which had long pauses and few intertextual connections[72] between them. The *ancient phase* had its most intense period in the fourth and third century BCE, but it continued long after that.[73] The *medieval phase* started in the twelfth century, had an especially fruitful period during the fourteenth century, and continued on beyond the fifteenth century.[74] And our ongoing *modern phase* of discussing the Liar paradox started only at the end of the 19th century.[75]

The ancient history of the Liar paradox begins with its invention – or discovery[76] –, which is generally attributed to Eubulides the Megarian.[77] While Plato did

69 Geach 1955; Prior 1955.

70 Cf., e. g., Beall 2009 and the literature mentioned there.

71 Cf. Read 1979.

72 Paul Vincent Spade and Stephen Read show that there have been surprisingly few links between the ancient and the medieval phase (Spade/Read 2009, section 1). And although there is a considerable amount of contemporary interest in medieval approaches to the Liar paradox, this must be seen as a *re*-discovery, because (to anticipate what we will see at the end of this section) the roots of modern renewed interest in the Liar paradox do not lie in a philosophical tradition but in new developments in mathematics that occurred at the end of the 19th century.

73 Spade/Read 2009, section 1. Cf. Bocheński 1956, 150–153, as well as Künne 2013, 119ff.

74 Spade/Read 2009. Cf. Bocheński 1956, 277–292 and Spade 1975.

75 Much of the pre-modern material on the Liar paradox is collected in Rüstow 1910, Bocheński 1956, and Spade 1975. (This material (as usual) belongs solely to European philosophy; research into Arab treatments of the Liar paradox is only beginning; cf. Alwisha/Sanson 2009.) Overviews of the medieval period are provided by Dutilh Novaes 2008, Spade/Read 2009, and some passages in Sorensen 2005.

76 While even someone with only slight platonist inclinations will hold that a *proof* is out there to be discovered, it must be doubtful that a *paradox*, although it also is an argument, can be accorded more independence from thinking beings than the artifacts they create. The reason for this is the important part played by *appearances* for the existence of a paradox (cf. section 1.1).

77 Bocheński 1956, 151; Spade/Read 2009, section 1.

not discuss it,[78] Aristotle proposed a solution,[79] and Chrysippus and others are reported to have produced copious writings about it (most of which are lost).[80] But in spite of this productivity, and *pace* Philetas of Kos (who according to an often told anecdote worried so much about the Liar paradox that he lost his night's sleep and in the end also his life),[81] it seems that in general no great *systematic* relevance was attributed to the Liar paradox even during the heyday of ancient discussions. Later on, it was seen as no more than a curiosity, meriting only brief remarks from Latin authors like Cicero and Seneca.[82]

Medieval European philosophers discussed the Liar paradox extensively as one of a group of problems they called "insolubilia" but by no means deemed to be unsolvable[83]. They produced an amazing plurality of approaches. Most of these can be seen as precursors of contemporary approaches;[84] and some are of great systematic interest from a contemporary perspective: Arthur Prior turned to John Buridan's approach for inspiration,[85] and Stephen Read has done much to articulate and defend the subtle approach of Thomas Bradwardine.[86] But the huge amount of work medieval philosophers spent on the Liar paradox should not be seen as a sign that they were greatly troubled by it. This is pointed out by Paul Vincent Spade:

> "the medievals did not seem to have had any 'crisis mentality' about these paradoxes. Although they wrote a great deal about them, there is no hint that they thought the paradoxes were crucial test cases against which their whole logic and semantics might fail. [...] the medievals did not draw great theoretical lessons from the insolubles. They did not seem to think the paradoxes showed anything very deep or important about the nature of language and its expressive capacity".[87]

And Catarina Dutilh Novaes, who otherwise is more cautious in her assessment than Spade,[88] observes:

> "As Barwise and Etchemendy put it, 'the significance of a paradox is never the paradox itself, but what it is a symptom of' [...]; the medieval authors, by contrast, were mostly interested in the paradoxes themselves, as particularly difficult logical puzzles."[89]

An even greater air of unconcern characterizes two discussions of the Liar paradox that were produced in the mid 19th century, by Bernard Bolzano and Charles

---

78 Bocheński 1956, 151.

79 Aristotle, *Sophistical Refutations*, 25, 180a27-b7. Cf. Brendel 1992, 23f.

80 Bocheński 1956, 151; Spade/Read 2009, section 1.

81 Bocheński 1956, 151; Spade/Read 2009, section 1.

82 Cicero, e. g., does little more than pose the question:

> "A man says that he is lying. Is what he says true or false?" (Cicero, *Prior Analytics*, II, 96)

Cf. Sorensen 2003, 88f.

83 Spade/Read 2009, section 5.

84 Cf. Simmons 1993, 83ff. and especially Dutilh Novaes 2008.

85 Prior 1976. Cf. Müller 2002, 137–140.

86 Read 2002, 2006, 2008, and 2009; and cf. the other contributions in Rahman/Tulenheimo/Genot 2008.

87 Spade 1982, 253.

88 Dutilh Novaes 2008, 228.

89 Dutilh Novaes 2008, 228; cf. Barwise/Etchemendy 1987, 4.

Sanders Peirce. Neither is longer than three pages, and Bolzano concludes his piece: "But enough of this sophistry!"[90] The interest of both modern philosophers seems to have been kindled by some medieval treatise on insolubles; Bolzano reacts to Savonarola and Peirce to Paul of Venice. Their contributions are probably nothing more than late echoes of the medieval debate; Bolzano and Peirce are neither proponents nor precursors of the modern discussion about the Liar paradox.

The modern phase in the history of the Liar paradox begins in the late 19th century with Georg Cantor's invention of set theory and in particular with his diagonal argument, which is a method of proof that led Bertrand Russell to discover his famous paradox of the set of all sets that are not an element of themselves.[91] Soon, further paradoxes were discovered – some of them are like Russell's paradox insofar as they involve set theoretical notions, others are like the Liar paradox insofar as they involve semantic notions. In the early years of the 20th century, the set theoretic paradoxes were recognized as a serious threat to the newly laid foundations of mathematics;[92] and in the 1930s, Alfred Tarski showed that the Liar paradox could be a serious threat to the new formal science of semantics.[93] These foundational problems explain the modern "crisis mentality"; and in their wake there was new interest in the Liar paradox[94] and its history.[95]

But the connection between Cantor's diagonal argument and the Liar paradox is not only historical but also systematic. The diagonal argument not only caused renewed interest in the Liar paradox, it also bears a resemblance to it because it involves a construction that is structurally similar to a Liar sentence. We will study this structural similarity in the next section (1.7), and then we will turn to the larger group of paradoxes which the Liar paradox belongs to (in section 1.8).

1.7 Set theory and Cantor's diagonal argument

Cantor used his diagonal argument to prove some important theorems of set theory which concern the size of sets with infinitely many elements. But before we come to his proof, I would like to make some brief and opinionated remarks about the notion of a set, about set theory, and about set theoretical parlance. The opinionated part does not concern set theory itself (which as a mathematical subject is beyond philosophical reproach), but the practice common among analytic philosophers to try and use set theoretical parlance as a kind of *lingua franca* and the corresponding inflationary use of the notion of a set in analytic metaphysics.

---

90 Bolzano 1978, 24–26; Peirce *Collected Papers* 5.340; my translation. The original text is: "Doch schon genug von dieser Spitzfindigkeit!" Bolzano's discussion of the Liar paradox is part of a letter he wrote in 1848; Peirce's discussion was originally published in 1868. Bolzano's approach is discussed in Brendel 1992, 42 and Künne 2013, 71ff.

91 Russell 1903, 101.

92 Cf., e. g., Deiser 2002, 183ff; Ebbinghaus 1994, 7ff.; Tiles 1989, 114ff.; Potter 2004, 25ff.; and especially Rheinwald 1988, 282ff.

93 Cf. sections 3.6 and 3.7.

94 Russell 1967[1908], 153ff.; Whitehead/Russell 1910, 63ff.; and cf. Russell 1985[1959], 59ff.

95 Cf., e. g., Rüstow 1910.

It is important that we distinguish *set theory*, which is part of modern mathematics, and the metaphysical notion of a set, which is among its origins (at least historically). The metaphysical notion entails that a set is a collection of objects that is itself conceived as an object. According to Cantor, it collects these objects "into a whole", and according to Felix Hausdorff, it collects them "into a new object".[96] To see which metaphysical commitments are connected to this notion of a set, it is helpful to contrast it with the notion of a *concept* and the notion of a *plurality*. In contrast to a concept, of which we may well want to deny that it is a special kind of object,[97] a set clearly is an *object*.[98] And a set is a *single* object, in contrast to a plurality like a flock of birds (of which we say '*they are* flying' rather than '*it is* flying'). By talking of the set of some objects one thus commits to the claim that for the collection of objects in question there exists an object, namely the set which has just those objects as its elements.

It is this metaphysical notion of a set that was at the origin of mathematical set theories,[99] which were formulated in an informal way first and foremost by Cantor (but also by Richard Dedekind and others) and later given axiomatic treatments by Ernst Zermelo, Abraham Fraenkel, John von Neumann, and others.[100] In mathematics, set theory plays a dual role, because it both is a subdiscipline among other mathematical subjects (distinct, e. g., from geometry, analysis, and algebra), and its ontology and language are used often as a background metaphysics and a *lingua franca* for the formulation of all of mathematics (now including the subdisciplines of geometry, analysis, and algebra).[101]

Set theoretical parlance has become a widespread dialect also in analytic philosophy, where some people would not shy away from construing a flock of birds as a (flying) set. Aside from the scientific glamour associated with set theory, this may well be due to reservations regarding pluralities. But although a majority of analytic philosophers view pluralities with suspicion, there are some who have made good use of the concept in important contexts.[102]

In his rant "Against Set Theory", Peter Simons reports the misuse of set theory in philosophy and speaks forcefully against it.[103] However, although I am sympathetic to his critique and prefer pluralities myself, I here do not want to commit myself in the matter. Instead, I will try to adopt a neutral jargon in this study: Aside from

---

96 Cantor 1895, quoted in Deiser 2002, 15, and Hausdorff 1914, 1.

97 Cf. section 4.2, where we will endorse the Fregean claim that concepts are not objects.

98 In addition to their objecthood, the *extensionality* of sets is a further feature which sets them off from concepts: A set x is identical to a set y if and only if they have the same elements, but on many construals concepts need not be equivalent if the same objects fall under them.

99 Thus even axiomatic set theory primarily is *not* to be understood in an *algebraic* way, i. e., as a formal theory which is about whatever satisfies its axioms, but in an *assertory* way, i. e., as being about a specific subject matter. Cf. Shapiro 2005, 67; and sections 3.1 and 4.2.

100 Cf., e. g., Ebbinghaus 1994, 7ff.

101 In recent years, category theory has been a new contender for this role; cf. Awodey 2010.

102 E. g., Peter Simons uses them to do justice to the metaphysics of ordinary objects, e. g., of the constitution of an orchestra out of musicians (Simons 1987, 144ff.) and George Boolos employs pluralities to give a non-standard interpretation of second order logic (cf. Boolos 1995, 54ff.).

103 Simons 2005.

set theory itself (which is of course about sets and their elements), we will, where others would speak *Settish*, talk about a *collection* and *the objects which are among it*. This talk can be spelled out *either* in terms of a set and its elements *or* in terms of a plurality and the many single objects which constitute it.

*

That being said, let us turn to some basic notions of the theory of sets which are needed to present Cantor's proof.[104] We write '$x \in y$' to express that the object x is an element of the set y. We say that a set x is a *subset* of a set y if and only if every object that is an element of x is an element of y. The set of all subsets of a set x is called the *powerset* of x; we write '$\mathcal{P}(x)$'.

To be able to speak about the size of infinite sets, we need to extrapolate our ordinary notion of the size of a collection, which concerns *finite* collections, to the notion of *cardinality*, which concerns not only finite but also infinite sets. Inspired by the everyday operation of counting, the sets x and y are said to *have the same cardinality* if and only if there is a bijective function from x to y (i. e., intuitively speaking, each element of x can be paired off with an element of y in such a way that no elements of x or y remain single).[105] A set x is said to be *of (properly) smaller cardinality* than a set y if and only if there is a bijection from x to a subset of y and there is no bijection from x to y.

Now, what Cantor showed with the diagonal argument – *Cantor's Theorem* – is that, even in the infinite case,[106] every set has a smaller cardinality than its powerset. Specifically, the set of natural numbers has a smaller cardinality than the set of sets of natural numbers. And as the technical notion of cardinality explicates and extends our intuitive notion of the size of a collection, we can say more simply that any set has more subsets than elements and that there are more sets of numbers than numbers.

In view of what Cantor's theorem says we should be able to distinguish different infinite cardinalities. We say that a set is *enumerable* if and only if it has the same cardinality as some subset of the natural numbers. Thus we can distinguish between sets which are *finite* (and hence enumerable), sets which are *enumerably infinite*, and sets which are *indenumerable*, i. e., infinite and not enumerable.[107]

The proof of Cantor's Theorem goes like this:

> Let f be an injective function from a set m to its powerset $\mathcal{P}$(m). We show that f is not surjective, i. e., that there is a $y \in \mathcal{P}$(m) such that there is no $x \in$ m with f(x) = y. For consider the set y which has as elements all and only those objects z

---

[104] Cf. Ebbinghaus 1994; Deiser 2002; and Smullyan/Fitting 2010 for set theory, and cf. Tiles 1989 and Potter 2004 for the history and philosophy of set theory.

[105] A function f from a set m to a set n is *surjective* if and only if for every $y \in n$ there is an $x \in m$ with f(x) = y; a function f is *injective* if and only if for every $x, y \in m$, if f(x) = f(y) then $x = y$; and a function f is *bijective* if and only if it is surjective and injective.

[106] In the infinite case, some of our ordinary intuitions about the size of collections fail. Notably, an infinite set is of the same cardinality as some of its proper subsets; e. g., there is a bijection between the natural numbers and the even numbers. Therefore Cantor's result is in fact more surprising than it might seem to someone acquainted only with finite collections.

[107] Cf., e. g., Smith 2007, 13f.

such that (i) $z \in$ m and (ii) $z \notin$ f(z). Then $y \in \mathcal{P}$(m) because of (i). But because of (ii) there can be no $x \in$ m such that f(x) = y. For assume (for reductio) that there is an x with f(x) = y. Is $x \in y$? If $x \in y$, then, because of f(x) = y and (ii), $x \notin y$. And if $x \notin y$, then, again because of f(x) = y and (ii), $x \in y$. Either way, there is a contradiction. Hence there is no such x; and therefore f is not surjective. Since m and f were arbitrary, we have shown that no injective function from a set to its powerset is surjective.[108]

But there was nothing obviously *diagonal* about that. To bring out what is diagonal about this method of proof, we will now turn to a special variant of the argument that concerns the set of natural numbers (i. e., of the numbers 0, 1, 2, 3, and so on). – Are there as many sets of natural numbers as there are natural numbers? A negative answer is already entailed by Cantor's Theorem, but it can also be justified more directly.

If there were as many sets of natural numbers as natural numbers, then we could give an (infinite) list of these sets like this: m_0, m_1, m_2, m_3, etc. And for any set of natural numbers m, we can specify its element by answering an (infinite) list of questions: Is $0 \in m$? Is $1 \in m$? Is $2 \in m$? Is $3 \in m$? Etc. In other words, we can specify any set of natural numbers by giving an infinite sequence of 'yes's and 'no's.[109] Thus we can specify any list of sets of natural numbers by an (infinite) table that has 'yes's and 'no's as its entries.

To give an example, let m_0 be the set of all odd numbers, m_1 be set that has zero as its only element, m_2 the set of all even numbers and m_3 the set of all prime numbers. We can present these bits of information in the following (infinite) table:

	Is $0 \in m$?	Is $1 \in m$?	Is $2 \in m$?	Is $3 \in m$?	etc.
m_0	**no**	yes	no	yes	...
m_1	yes	**no**	no	no	...
m_2	no	no	**yes**	no	...
m_3	no	no	yes	**yes**	...
etc.	...	...	...	...	**...**

Now we can specify a set that cannot occur at any position of our list. For take the *diagonal sequence* (here: 'no, no, yes, yes, ...') and *invert* it by changing each 'yes' into a 'no' and each 'no' into a 'yes'. The resulting *inverse-diagonal sequence* (here: 'yes, yes, no, no ...') specifies a set of natural numbers that cannot occur in the list of sets of natural numbers we started with. For assume that it occupies the n^{th} position in the list. Then the sequence of 'yes's and 'no's at its n^{th} place could neither have a 'yes' nor a 'no', on pain of contradiction. Therefore the set of sets of natural numbers is not listable – it is not of the same cardinality as the natural numbers.[110]

---

108 For expositions of the diagonal argument, cf. Boolos/Burgess/Jeffrey 2002, 16ff., or any other textbook introducing mathematical logic. Simmons 1993 presents a whole theory of diagonal arguments, already with a view to their similarity to the reasoning of the Liar paradox.

109 In technical jargon, this amounts to specifying a set of natural numbers via its *characteristic function*.

110 Note that this argument, which has an obviously diagonal character, is a special case of the proof we saw above for Cantor's theorem. *Listing* the sets m_0, m_1, m_2, m_3, etc. amounts to specifying a function

This presentation of the above proof of Cantor's Theorem does not only show what is diagonal about it, but it can also illustrate the resemblance between the diagonal argument and the Liar paradox. The resemblance concerns not the arguments themselves, but the objects reasoned about, because it is the specification of a set by the inverse-diagonal sequence that has a similar structure as a Liar sentence: Something *points back at itself in a negative way*. A Liar sentence points back at itself by way of reference, and obviously in a negative way because it ascribes falsity or untruth. This general characteristic of the Liar paradox has sometimes been noted. Bradwardine defines an insoluble as a certain kind of "difficult paralogism" that arises "from some [speech] act's *reflection on itself with a privative determination*";[111] Mackie says that a Liar sentence "depend[s] inversely on itself"[112]. Similarly, the inverse-diagonal sequence points back at itself insofar as it is a *diagonal* sequence, and in a negative way insofar as it is *inverse*.[113]

This shows how Cantor's diagonal argument is more than the historical origin of modern interest in the Liar paradox – it is systematically similar on a deep level, because Cantor transferred the structure of *privative reflection* into a mathematical setting.

1.8 Semantic and set theoretic paradoxes

The Liar paradox is one member of a larger group of paradoxes that are often seen as forming two different families, nowadays called the *semantic paradoxes* and the *set theoretic paradoxes*.[114] Here is a double list of important paradoxes belonging to the two families:

semantic paradoxes:	set theoretic paradoxes:
the Liar paradox	Russell's paradox
Curry's paradox	Cantor's paradox
Berry's paradox	Burali-Forti's paradox
Grelling's paradox	Richard's paradox
	Mirimanoff's paradox

from the natural numbers to sets of natural numbers, which corresponds to the function f. Inverting the diagonal corresponds to condition (ii). And the set specified by the inverse-diagonal sequence corresponds to the set *y*.

111 Spade 1975, 106; the English translation is from Spade/Read 2009, section 5; the addition "[speech]" is by Spade and Read; the italics are mine.

112 Mackie 1973, 256.

113 It is not the *set* specified by the inverse-diagonal sequence that points back at itself (sets do not point to anything), but rather the *sequence* itself. And the sequence does its back-pointing only in virtue of the particular listing we work with. This need not be seen as a structural difference to the Liar paradox, because the reflective nature of a Liar sentence is also grounded in its position within a larger framework: It can only refer to itself in virtue of which meanings attach to its terms.

114 The distinction was first made (in other terms) in Peano 1906 and Ramsey 1990[1925], 183f. Some overviews and discussions are: Kneale/Kneale 1962, 652–657; Rheinwald 1988, 14f.; Brendel 1992, 45–54; Priest 1994 and 2002, 141–143; and Deiser 2002, 183–194.

Before we turn to the question of what distinguishes the families of semantic and of set theoretic paradoxes, let us look briefly at the last two semantic paradoxes and the first two set theoretic paradoxes of the double list.[115]

Berry's paradox concerns a description like 'the least natural number that cannot be described in under thirteen words'. The paradoxical reasoning goes like this: There are only finitely many descriptions of a given length, but infinitely many natural numbers, so there must be natural numbers that cannot be described in under thirteen words. The least of them satisfies the description 'the least natural number that cannot be described in under thirteen words'. But then it *can* be described in under thirteen words; a contradiction.

Grelling's paradox concerns the predicate '… is heterological', which is satisfied by all and only those predicates that do no satisfy themselves. The predicate '… is heterological' seems to be unproblematic as long as it is applied to other predicates; we can say for instance that while 'monosyllabic' is heterological, 'polysyllabic' is not. But the predicate makes trouble when we try to apply it to itself; '… is heterological' is heterological if and only if '… is heterological' is not heterological; a contradiction.

Russell's paradox concerns the set of all sets that are not an element of themselves, r. The notion of not being an element of itself seems to be unproblematic as long as it is applied to other sets; we can say for instance that while the set of teacups is not an element of itself, the set of abstract objects is. But it makes trouble when it is applied to the set r itself. r's definition entails that every set must satisfy the following condition: $x \in \mathrm{r}$ if and only if $x \notin x$. But as r is a set, this condition must also hold of r. Thus, $\mathrm{r} \in \mathrm{r}$ if and only if $\mathrm{r} \notin \mathrm{r}$; a contradiction.

Cantor's paradox concerns the set of all sets, v. Due to v's containing *all* sets, every subset of v must be an element of v, so that v is its own powerset. Because of Cantor's Theorem, v does not have the same cardinality as itself – but every set has the same cardinality as itself. A contradiction.

The set theoretic paradoxes do not occur in the systems of set theory that were axiomatized in the first third of the 20$^{\text{th}}$ century (which are what will be found in a standard textbook on the mathematical subdiscipline of set theory).[116] Their place is in *naïve set theory*, which is a theory of sets based on the assumption, often called the *naïve comprehension principle*, that for *every* concept there is a set of all and only those objects that fall under that concept. If, in particular, we use some basic concept of set theory itself to define a set, paradox is wont to follow. For example, Russell's paradox is based on the notion of *elementhood* and Cantor's paradox is based on the notion of *sethood*.

It is usually said that what distinguishes the two families of paradoxes is that, besides logic, set theoretic paradoxes involve only set theoretic notions and semantic

[115] The Liar paradox is of course discussed in enough places of this study, and Curry's paradox will be presented in section 1.5. Burali-Forti's, Richard's, and Mirimanoff's paradox involve notions that are too technical for a presentation here.

[116] Two important axiomatic set theories which have shown no sign of paradox so far are *Zermelo-Fraenkel set theory* (ZF or ZFC) and the *Neumann-Bernays-Gödel theory of sets and classes* (NBG); cf. Ebbinghaus 1994, especially 7–14; Deiser 2002, 183ff.; and Smullyan/Fitting 2010, especially 11–14.

paradoxes involve only semantic notions, i. e., notions like truth, satisfaction, reference, and description.[117] And to be sure, the Liar paradox involves the notion of *truth*, Grelling's paradox the notion of *satisfaction*, and Berry's paradox the notion of *description*. But there has been considerable dispute about whether this criterion provides a justification for seeing these paradoxes as forming two distinct and essentially different families.

An historical reason for seeing an essential difference is that there is a consensus among set theoreticians that the set theoretic paradoxes are solved, while there is no such consensus among the people who work on semantics (and especially not among those who work on the theory of truth) about a solution to the semantic paradoxes. The main systematic reason for seeing an essential difference is, naturally, that the notions appealed to in the above criterion belong to essentially different disciplines, set theory and semantics.[118] But this has been contested. Graham Priest, e. g., points out that after the formalization of considerable parts of semantics it has become difficult to disentangle as a discipline from set theory.[119] And I would like to argue that, conversely, set theory is difficult to disentangle from semantics, because the naïve comprehension principle postulates the existence of a set as the extension of every *meaningful predicate*, and the meaningfulness of predicates of course falls into the province of semantics. A further systematic reason for the underlying similarity between the paradoxes of both families, by contrast, is positive. For the paradoxes presented above, it is not too difficult to see a common structure, namely the *structure of privative reflection* that was described in the last section (1.7).[120] And every one of the paradoxes in the above double list at least involves a structure of *reflection*, or, to use Russell's term, "circularity";[121] therefore they and their relatives are sometimes called "the paradoxes of self-reference".[122]

---

[117] Tarski 1956[1936], 401; Tarski 1944, 345. Cf. Künne 2003, 176–180; Sher 2005, 150. – Roughly, a notion is semantic in this (Tarskian) sense if and only if it is in some way about a connection between a language and the world represented by it. Cf. section 4.1 for more on why these notions are called *semantic*.

[118] Elke Brendel sees *the problem of expressibility* as distinctive of the semantic paradoxes, thus giving a further systematic reason for seeing an essential difference to set theoretic paradoxes (cf. Brendel 1992, 45–54). The problem of expressibility is, to anticipate section 2.8, the problem that key notions employed in a solution to a paradox turn out to be inexpressible in the paradox-free language or theory the paradox-solver tries to develop. But there are also problems of expressibility for important solutions to the set theoretic paradoxes. Frederic Brenton Fitch has pointed out that the sentences of Russell's theory of types themselves violate type restrictions (Fitch 1964); and Mirimanoff's paradox concerns the set of all well-founded sets, which is well-founded if and only if it is not well-founded (Deiser 2002, 185ff.), where *well-foundedness* is a key notion in all classical solutions to the set theoretic paradoxes (cf. Bromand 2001, 127ff.).

[119] Priest 2002, 142.

[120] This may be less obvious for Cantor's paradox as it is described here, but the structure of privative reflection also occurs in the diagonal construction that is used in the proof of Cantor's theorem (cf. section 1.7) which is an essential part of the reasoning behind Cantor's paradox.

[121] Whitehead/Russell 1910, 39ff.

[122] Cf., e. g., Priest 1994, 25 and Restall 1993, 279.

My own sympathies lie with the similarity view, but I cannot give a conclusive argument here.[123] The question of whether there are essential differences between semantic and set theoretic paradoxes is far too difficult to settle. An important reason for this difficulty is that probably no one can answer the question independently from his or her approach to these paradoxes.[124]

1.9 Just a joke?

People who are into neither logic nor analytic philosophy understand the reasoning of the Liar paradox easily, but in general they do not see it as a problem of any philosophical depth.[125] This phenomenon stands in an odd tension to the fact that the philosophical literature on the Liar paradox of the last one hundred years is truly vast.[126] So, how serious is the philosophical problem posed by the Liar paradox? Let us hear answers by contemporary philosophers who have worked on it, starting with two extreme positions.

Dorothy Grover is relaxed:

> "it is not easy to persuade others that we need to 'resolve the paradox of the liar'; or that analysis of the liar may reveal crucial insights. Non-philosophers may grant there is a curious puzzle, but it is a 'don't care' puzzle. How, and where, is the liar so crucial to our understanding? I will now argue that there is something right in the naïve reaction of unconcern."[127]

Vann McGee, on the other side of the spectrum, is dead serious:

> "There are scarcely any philosophical problems of greater urgency than the liar paradox, for there are scarcely any concepts more central to our philosophical understanding than the concept of truth. [...] The liar antinomy and the closely related antinomies involving reference show us, quite unmistakably, that our present way of thinking about truth and reference is inconsistent. Unless we can devise new ways of thinking about truth and reference we shall not have even the beginning of a satisfactory understanding of human language."[128]

In a similar vein, Jon Barwise and John Etchemendy move swiftly from considering a relaxed attitude like Grover's to adopting an attitude nearly as earnest as McGee's:

---

[123] But cf. Pleitz 2015a.

[124] On an historical note: Frank Plumpton Ramsey, whose 1925 text is the locus classicus for the distinction between semantic and set theoretic paradoxes (Ramsey 1990[1925], 183f.), drew attention to it in order to defend his favorite approach to the set theoretic paradoxes (i. e., Russell & Whitehead's) against the criticism that it did not solve the semantic paradoxes (Ramsey 1990[1925], 191).

[125] I was able to collect anecdotal evidence for this observation in conversations with friends and acquaintances about the topic of my research.

[126] This vastness is attested to, for instance, by the following four anthologies which appeared in the last decade: Beall 2003; Priest/Beall/Armour-Garb 2004; Beall/Armour-Garb 2005; Beall 2007a. Cf. also sections 7.4 and 7.5.

[127] Grover 2005, 177.

[128] McGee 1991, vii.

> "On first encounter, it's hard not to consider assertions of this sort [i. e., Liar sentences] as jokes, hardly matters of serious intellectual inquiry. But when one's subject matter involves the notion of truth in a central way, for example when studying the semantic properties of a language, the jokes take on a new air of seriousness: they become genuine paradoxes. And one of the important lessons of twentieth century science, in fields as diverse as set theory, physics, and semantics, is that paradox matters. [...] a paradox demonstrates that our understanding of some basic concept or cluster of concepts is crucially flawed, that the concepts break down in limiting cases. And although the limiting cases may strike us as odd or unlikely, or even amusing, the flaw itself is a feature of the concepts, not the limiting cases that bring it to the fore. If the concepts are important ones, this is no laughing matter."[129]

Why are the reactions of philosophers to the Liar paradox so diverse? I think that we will find some clues for an answer in a frank observation James Cargile makes about his fellow philosophers and logicians.

> "Semantic paradoxes are very commonly regarded by philosophers as trifling problems. Of course they are taken seriously by logicians, but most philosophers are not logicians, and to this majority, the paradoxes are ranked intellectually about on a par with party games or newspaper 'brain teasers'. Handed a card with 'The statement on the other side is false' on one side and 'The statement on the other side is true' on the other, even an intellectually serious person may consider an amused smile an appropriate response.
>
> One way of getting the problems treated as serious is to treat them mathematically. The introduction of formal symbolism tends to have a sobering effect, and since it is not generally understood, it makes it less embarrassing to be overheard discussing these problems."[130]

If Cargile is right, taking the Liar paradox seriously is somehow connected to taking a mathematical or formal[131] approach in philosophy. And indeed, while philosophers of all persuasions think of the notion of truth as important,[132] the philosophers who are troubled that the notion of truth might be compromised by the consequences of the Liar paradox tend to be just those who like to incorporate mathematical or formal methods in their work.[133]

---

[129] Barwise/Etchemendy 1987, 4.

[130] Cargile 1979, 235.

[131] The notion of formality employed here is basically the vulgar one according to which a theory is formal if its formulation contains lots of formal symbols. The vulgar notion of formality works well enough in describing the day-to-day interactions of philosophers. Other notions of formality, which are more important systematically, can be derived from the concept of a formal language that we will characterize in section 3.1.

[132] A case in point is Wolfgang Künne, who in his monumental monograph about theories of truth says about the debate about the Liar paradox in the late 20th century: "As I had to confess already in the preface to this book, I have nothing enlightening to say about, let alone to contribute to this debate. So I quickly, and somewhat shamefacedly, move on [...]" (Künne 2003, 203). Ten years later, however, he did present a (less monumental) study on the Liar paradox (Künne 2013).

[133] To anticipate, this claim can be substantiated by noting that the proponents of *the meaninglessness solution* adopt the no problem attitude and in general decline using formal methods while the proponents

Let us tentatively (and only for the descriptive purposes of the present section) distinguish between *formal* and *non-formal* philosophers. Of course, on a systematic level any distinction between formal and non-formal *philosophy* would be difficult to uphold. But we are concerned here only with the attitudes contemporary working philosophers have to their everyday work. And there a formal and a non-formal fraction can indeed be made out – perhaps a late echo of the division of early 20th century analytic philosophers into philosopher-logicians (like the earlier Wittgenstein) and ordinary language philosophers (like the later Wittgenstein), perhaps no more than a reflection of personal tastes and talents.

In these terms we can say, roughly, that where non-formal philosophers say 'no problem', formal philosophers cry 'no joke!' And usually, they do not listen much to each other. The lack of communication between the two fractions is another important aspect of the situation characterized by Cargile. Volker Halbach describes the situation in the case of philosophical theories of truth:

> "When we look at the literature by analytic philosophers on the topic of truth, it is difficult to resist the impression, not only that very heterogeneous projects and programs are pursued under the same label, but also that two completely independent areas of research have developed. Each has its own key literature, its own prominent and trend-setting protagonists, its own school and paradigms. Mostly, the people working in one area will not quote the contributions of the other area and – so we can surmise – will not even be aware of them."[134]

So the diversity of philosophical reactions to the Liar paradox can plausibly be explained by the nearly conflict-free co-existence of two camps of philosophers, which is nearly conflict-free because intellectual conflict presupposes communication.

1.10 The question of formality

Let us not contribute further to this sorry state of affairs. I propose that we adopt and if possible defend the *conciliatory hypothesis* that probably there is some justification *both* for the no-joke *and* for the no-problem attitude. In order to do justice to both sides I set myself a twofold aim in writing this study: First, in parts I and II (especially in chapters 2 through 5), I want to convince non-formal philosophers (and everyone else) that the Liar paradox is indeed a serious problem. But then, in part III, I want to convince formal philosophers (and everyone else) that the Liar paradox in the end of the day will turn out to have been not much more than a joke, by proposing a solution that is meant to literally explain *away* the whole problem.

Because of the (loose) correlation we found in the preceding section between taking the Liar paradox seriously and belonging to the camp of formal philosophers, we will also present the Liar reasoning in a non-standard way in the following two chapters: They are devoted to a presentation of the basic Liar reasoning which deals

---

of what we will call *approaches of sophisticated surgery* and *approaches of palliative care* adopt the no joke attitude and enjoy using formal methods. Cf. section 7.5.

[134] Halbach 2005, 229; my translation.

with informal and formal variants *separately*. Our discussion will be guided by what can be called *the question of formality*, which we can pose here only in the rough form of the following collection of sub-questions:

How formal are the informal variants of the Liar paradox?

How informal are its formal variants?

Are the formal variants more serious than the informal ones, or vice versa?

What is formality, anyway?

In chapter 2 we will present the basic Liar reasoning in an informal way; in chapter 3 we will turn to the notion of formality and present the basic Liar reasoning in a formal way. The methodological device of splitting up informal and formal variants of the Liar paradox upon two chapters while keeping an eye on the guiding question of formality will enable us to understand better the serious problem posed by the Liar paradox, and to see that its seriousness is not bound up to one particular way of doing philosophy.

Chapter 2 Informal Logic and the Liar Reasoning

Arthur Prior:

"Logic is commonly thought of as having something to do with argument, in fact as being the systematic discrimination of good arguments from bad; and, as a first approximation, this will do."[135]

There are many good introductions to the Liar paradox.[136] We can therefore allow ourselves to give a non-standard presentation of the Liar reasoning, splitting it up into a non-formal presentation in this chapter and a formal presentation in the next chapter. This will not only allow us to get a deeper understanding of the seriousness of the problem posed by the Liar paradox,[137] contrasting the formal approach to logic which by now is common to an informal approach that is no less warranted (in sections 2.1 and 3.1) and comparing different ways of explicating basic intuitions about truth (in sections 2.2 and 3.2, and in sections 2.3 and 3.3) will also help us answer the question of formality.[138] Along the way we will establish some terminology and introduce some notions which will be important later on.

The present chapter starts with a list of some rules and laws of informal logic (in section 2.1), followed by a formulation of some principles about truth and falsity that are essential to the basic Liar reasoning (in sections 2.2 and 2.3). The central part of the chapter is a comparison of a number of informal variants of the basic Liar reasoning (in section 2.4), which will show that the paradox does by no means presuppose formal logic and is not connected essentially to classical logic. After that we will look at how the Liar reasoning typically goes on to unfold after a first attempt has been made to thwart the problem (in sections 2.5 through 2.10). Although this presentation of the extended Liar reasoning plays no part in the contrast between informal and formal variants of the Liar paradox that is the overall aim of chapters 2 and 3, it is best given at the end of the present chapter because an informal presentation of these matters will suffice for our purposes.

---

135 Prior 1962[1955], 1.

136 Introductions to the Liar paradox can be found as single contributions (Quine 1976a; Visser 2004; Dowden 2010; Beall/Glanzberg 2014), as chapters in books which introduce the philosophy of logic (Haack 1978, 135–151; Read 1995, 148–172; Priest 2000, 31–37), in books which introduce the theory of truth (Soames 1999; Burgess/Burgess 2011), as chapters of books which are devoted to a wider range of paradoxes (Cook 2001, 30–61; Sainsbury 1988, 114ff.; Rescher 2001, 199ff.; Priest 2002, 141–155; Sorensen 2005 (cf. the entry in the index on page 387); Clark 2007, 112–119), as introductory chapters in anthologies devoted to the Liar paradox (Martin 1984b; Beall 2007b), and as introductory chapters of monographs specialized on the Liar paradox or a larger group of paradoxes (Barwise/Etchemendy 1987, 3–25; Rheinwald 1988, 9–56; McGee 1991; Brendel 1992, 3–17; Simmons 1993, 1–19; Gupta/Belnap 1993, 1–32; Bromand 2001, 11–36; Priest 2006a, 19–27; Field 2008, 1–19).

137 Cf. sections 1.9 and 1.10.

138 Cf. the end of section 1.10.

As we have already given one informal presentation of the Liar paradox in the first chapter,[139] a word of explanation is in order before we set out to give a second one. The two informal presentations differ in their aim and are aimed at a different (imagined) audience. In the previous chapter, our aim was to introduce the Liar paradox and its main ingredients in a thoroughly unchallenging way, without going into any depth when describing those ingredients.[140] Although the actual readers of this study will likely be familiar with logic, the presentation in the introductory first chapter was directed at an imagined audience of un-initiated readers. In the present chapter 2, in contrast, our (official) aim is to convince an audience that is familiar with the usual more formal ways of presenting the Liar reasoning that all of that reasoning can already be done in an informal way, as well as to remind this audience of how robust the Liar reasoning is.[141] Thus there is no redundancy in giving another informal presentation.

2.1 Informal logic

Logic is concerned with what makes an argument valid, i. e., with the rules and laws of good reasoning. Formal logic explicates good reasoning by giving a complete specification of a formal language and a detailed description of the inferential relations between sentences of that formal language.[142] By *informal logic* we shall mean a specification of good reasoning from *within* the language we actually reason in (or in a language that is akin to the language we reason in). That language is not formal, at least not in the sense of formal logic that will be explained in the next chapter.[143] But it may well be *regimented*, for instance by the stipulation that the word 'and' is to be used unambiguously as connecting sentences in a way that disregards their order as well as any temporal or causal connotations. And it may even contain *formal symbols*, but these will be nothing more than abbreviations of logical expressions of our language (usually regimented ones). E. g., in the symbolically augmented English we use in this study,

'$=$' abbreviates 'is identical to',

'$=_{\text{def}}$' abbreviates 'is by definition identical to',

'$\Rightarrow$' abbreviates 'if … then', and

'$\Leftrightarrow$' abbreviates 'if and only if'.

Here is a list of rules of informal logic that will play a role later:[144]

[139] Cf. section 1.2.

[140] The logic ingredients will be discussed in chapters 2 and 3, and the language ingredient in parts II and III.

[141] The occasional reader who is truly uninitiated but interested is of course invited to read on! Most of the more difficult material will be explained in a self-contained way.

[142] Cf. chapter 3 and in particular sections 3.1 through 3.3.

[143] Cf. section 3.1.

[144] We present these rules of informal logic in a non-standard way here, phrased in a temporalized way and not mentioning but using sentences. Further, to distinguish informal from formal logic, we say

conjunction introduction:	If (at some step of our reasoning) we have inferred *that p* and *that q*, then we can (go on to) infer *that p and q*.
modus ponens:	If we have inferred *that if p then q* and we have inferred *that p*, then we can infer *that q*.
conditional proof:	If from the assumption *that p* we have inferred *that q*, we can infer *that if p then q*.
transitivity of 'if and only if':	If we have inferred *that p if and only if q* and *that q if and only if r*, then we can infer *that p if and only if r*.
substitution of equivalents:	If we have inferred *that p if and only if q* and we have inferred *something that includes the claim that p*, then we can substitute the claim *that q* for the claim *that p* in it.
reasoning by cases:	If we have inferred *that p or q*, *that if p then r*, and *that if q then r*, then we can infer *that r*.
consequentia mirabilis:	If from the assumption *that p* we have inferred *that it is not the case that p*, then we can infer *that it is not the case that p*.
double negation elimination:	If we have inferred *that it is not the case that it is not the case that p*, then we can infer *that p*.
reductio ad absurdum:	If from the assumption *that it is not the case that p* we can infer a contradiction (e. g., *that q and it is not the case that q*), then we can infer *that p*.
ex falso sequitur quodlibet:	If we have inferred *that p and it is not the case that p*, then we can infer *that q* (for any arbitrary claim *that q*).
De Morgan:	If we have inferred *that it is not the case that p* and *that it is not the case that q*, then we can infer *that it is not the case that p or that q*.

A logical law can be conceived as the limit case of a logical rule; that is, as a logical rule without any precondition of the form 'if we have inferred …'. The following two logical laws play important roles in connection with the Liar paradox:[145]

(Excluded Middle)	We can infer *that p or it is not the case that p*.
(Non-Contradiction)	We can infer *that it is not the case that (p and it is not the case that p)*.

'infer' instead of 'deduce' or 'derive', 'logical rule' instead of 'inference rule', and 'logical law' instead of 'theorem' or 'valid formula'. Some readers might nevertheless be reminded of the inferential rules of a natural deduction variant of the deductive system of classical logic (cf. section 3.1). But that is due mainly to the success of natural deduction at emulating real reasoning. There are also important differences, because we here do not give a complete list of rules meant to specify the meanings of the connectives, some of our rules would have the status of derived rules in a natural deduction system, and all that only modulo formalization, i. e., the translation into a formal language.

145 Cf. section 2.3.

Each one of these rules (and laws) of informal logic corresponds to an inference rule (or theorem) of classical propositional logic.[146] And that is as it should be. Classical propositional logic is, after all, considered by many to be (an important part of) a codification of our best reasoning – in other words, it is a promising candidate for (an important part of) an explication of informal logic. Others disagree, for a variety of reasons.[147] However, we need not take a stand on this matter here. What is important for our presentation of the Liar paradox is only that rejecting one or more of these rules amounts to departing from classical logic.

2.2 The naïve truth principle

The principles of truth and falsity that are appealed to in the basic Liar reasoning are of two kinds. Some are about the relation of truths and falsehoods to what is the case (and will be discussed in this section); others are about the relation between truth and falsity (and will be discussed in the next section).

The principles about how truths and falsehoods relate to what is the case that play a role in the Liar paradox are often traced back to Aristotle's famous statement about truth and falsity:[148]

> "To say of what is that it is not, or of what is not that it is, is false, while to say of what is that it is, and of what is not that it is not, is true."[149]

This is open to different interpretations, especially as it is not clear what "what is" and "what is not" mean. But very probably Aristotle's dictum is meant to sanction claims like the following: To say that Socrates is famous is true, because Socrates is famous. As Plato has not been forgotten, it would be false to say that Plato has been forgotten. And so on.

A formulation that sounds more precise (at least to modern ears) is given by Alfred Tarski:

> "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".[150]

This is easily supplemented[151] by a corresponding principle about falsity:

---

146 Cf. section 3.3.

147 For overviews of non-classical logic, cf. Haack 1978 and Priest 2008.

148 Less famously, Plato had already formulated a similar principle: "a true proposition says that which is, and a false proposition says that which is not" (Plato, *Cratylus* 385b; cf. *Sophistes* 240e 10–241 a1).

149 Aristotle *Metaphysics*, Γ 7: 1011b 26–7. Künne aptly characterizes this English version of the famous principle of Aristotle as "stunningly monosyllabic" (Künne 2003; 95).

150 Tarski 1956[1935], 155.

151 It is common to use the second half of Aristotle's famous statement as an inspiration for a principle about how a truth relates to what is case. But the fact that its first half provides an equally good inspiration for a corresponding principle about how a falsehood relates to what is the case is often overlooked. Graham Priest, for example, defines falsity not directly as a mismatch with what is the case, but indirectly as truth of negation (Priest 2006a, 64). Thus he (knowingly) bars the way to a non-circular definition of negation in terms of truth and falsity: "It would seem that falsity and negation can