

THE APRIORITY OF LOGICAL TRUTH

Indrek Lobus

A Thesis Submitted for the Degree of PhD
at the
University of St Andrews



2023

Full metadata for this thesis is available in
St Andrews Research Repository
at:

<http://research-repository.st-andrews.ac.uk/>

Identifiers to use to cite or link to this thesis:

DOI: <https://doi.org/10.17630/sta/308>
<http://hdl.handle.net/10023/27072>

This item is protected by original copyright

This item is licensed under a
Creative Commons License

<https://creativecommons.org/licenses/by-nc-sa/4.0>

The Apriority of Logical Truth

Indrek Lobus



University of
St Andrews

This thesis is submitted in partial fulfilment for the degree of
Doctor of Philosophy (PhD)
at the University of St Andrews

October 2021

Printed copy

No embargo on print copy.

Electronic copy

No embargo on electronic copy.

18 October 2021

Date

Indrek Lobus

18 October 2021

Date

Professor Peter Sullivan, FRSE

Underpinning Research Data or Digital Outputs

Candidate's declaration

I, Indrek Lobus, hereby certify that no requirements to deposit original research data or digital outputs apply to this thesis and that, where appropriate, secondary data used have been referenced in the full text of my thesis.

18 October 2021

Date

Indrek Lobus

Abstract

THE APRIORITY OF LOGICAL TRUTH

by

Indrek Lõbus

Principal supervisor: Peter Sullivan

Second supervisor: Colin Johnston

I defend the claim that knowledge of logical truths does not depend on experience, that it is a priori knowledge. In chapter I, I introduce and clarify the idea of a priori knowledge, and in chapter II, I respond to a widespread objection to it derived from Quine's critique of analyticity. In chapter III, I distinguish between two approaches to developing an account of a priori knowledge, one that ties apriority to the cognitive abilities of knowers and another that ties it to features of truths known. I motivate the second approach over the first. In chapter IV, I sketch an outline of how the second approach could lead to an account of the apriority of logical knowledge. In the final three chapters, I develop an account which follows the outline.

In chapter V, I introduce the Fregean idea of logic as a study of the laws of truth which I spell out with the help of the semantic conception of truth developed by Tarski. In chapter VI, I defend an idea found in Wittgenstein's *Tractatus* that the signs for logical operations cannot be construed as having objects corresponding to them. The defence proceeds via a defence of Frege's Context Principle. A consequence of the Tractarian idea is that logical truths are not expressions of thought, that they are truths without a truth-condition. In the last chapter, I spell out the epistemic implications of this idea, that an understanding of truth is accompanied by knowledge of every logical truth and that any acquisition of knowledge by experience therefore presupposes logical knowledge, meaning that logical knowledge itself cannot depend on experience or any other means of learning.

*By that which is straight we discern both straight and crooked;
for the carpenter's rule is the test of both,
but the crooked tests neither itself nor the straight.*

—Aristotle, *De Anima*, I. v. 411a5-7

Acknowledgements

I could not have written this thesis alone. I wish to thank my principal supervisor Peter Sullivan for all the help and support and the countless hours he spent patiently going through my chaotic notes. The thesis that I ended up writing is a tangible outcome of the research that I did under Peter's supervision, but I realize now that it was never the goal of that research. Not really. What Peter wanted me to take from my studies, I think, was an appreciation of the landscape through which he guided me, a landscape that he himself finds so clearly beautiful, and that I be able to find my way around it to explore it on my own. Thank you for introducing me to that world that I now feel at home in, and for being patient with me even at times when I must have seemed hopelessly lost. And thank you, most of all, for inspiring me to be a better philosopher and, dare I say, a better person.

I would also like to thank my second supervisor Colin Johnston for the support and the much-needed feedback that helped me avoid some terrible dead ends, and for helping me understand parts of the *Tractatus* that I previously did not know I did not understand. I wish to thank my annual reviewers at St. Andrews, especially Derek Ball and Peter Milne who oversaw my progress throughout most of my four years of study and gave helpful feedback and encouragements. And I wish to thank my bachelor's and master's thesis supervisor Daniel Cohnitz for helping me get this far.

I did my research at the University of Stirling as a member of the KBNS project. We met twice a week for three years to discuss topics related to a priori knowledge and self-knowledge. The discussions with fellow PhD students and other project members, both in the seminar room and outside it, had a critical role to play in the thesis receiving the shape that it now has.

I wish to thank Crispin Wright for organizing and leading the project, for bringing us all together, and for giving me and all of us the opportunity to do our research in an intellectually stimulating environment. I also wish to thank the other project members and everyone else who took part in our meetings: Giovanni Merlo, Adrian Haddock, Carrie Ichikawa Jenkins, Jonathan Jenkins Ichikawa, Sonia Roca-Royes, Alan Millar, Joshua Thorpe, Alisa Mandrigin, Michael Wheeler, Kent Hurtig, Philip Ebert, and especially Giacomo Melis for his support and encouragement. All of you contributed to the completion of this work either directly by your questions and feedback or by setting an example for how to approach the matter. I could point to passages in the thesis and say: this is a response to that question that Jonathan asked, that is

inspired by that comment made by Giovanni in the seminar, that by the example that Crispin gave, here I kept thinking how would Carrie respond, etc. In this way, your influence is present throughout this thesis and will surely continue to be present in my future work. Thank you.

I wish to thank my fellow PhD students many of whom also took part in the project meetings but more importantly with whom I shared an office space which was always a place for lively and inspiring discussions: Ásgeir Matthíasson, José Mestre, Xintong Wei, Jimena Clavel Vázquez, Moritz Baron, Sam Symons, and especially Paul Conlan. Our heated discussions with Paul which started at the office or the seminar often continued on the bus on our way back to town. I dare not imagine how we must have come across to the other passengers. Thank you for the support, the stimulating conversations, and the friendship.

Most of all, I wish to thank my best friend and fiancée Merike Reiljan for her love and support. Forgive me for putting you and both of us through this. Thank you for not letting me disappear.

Funding

The research was undertaken with the support of a studentship in the Knowledge beyond Natural Science research project, coordinated by Professor Crispin Wright and funded by the John Templeton Foundation and the University of Stirling.

Table of Contents

I. Introduction: A Priori Knowledge.....	1
1. The <i>Meno</i> and the Knowledge Within	1
2. A Priori and A Posteriori Knowledge	3
4. Justification by Experience	6
5. Why A Priori?	12
6. The Aim and Structure of the Thesis	14
II. Logic and the Web of Belief	18
1. Quine’s Confirmational Holism.....	18
2. A Problem for A Priori Knowledge?	23
3. Empiricism by Abduction?	27
III. Is A Priori a Way of Knowing?	32
1. A Priori as a Way of Knowing	33
2. A Priori as a Kind of Truth	39
3. Williamson’s Insignificance Challenge	47
4. The Insignificance of Ways of Knowing	51
IV. Alice on Logical Knowledge	58
1. Alice’s Thesis.....	59
2. A Limit to Epistemic Possibilities	62
3. Stalnaker’s Limit by Intension?	64
4. ‘Logically False Belief’ as Nonsense?.....	67
5. Truth-Values without Truth-Conditions	72
6. Two Kinds of Knowledge	73
7. Knowledge by Alice’s Thesis?	74
V. The Laws of Truth	78
1. Convention T	79

2. Tarski's Truth by Satisfaction.....	83
3. Thoughts	88
4. Frege on Generality.....	93
5. A Definition of Sentence	99
6. Atomic Propositions.....	102
7. A Definition of Truth	105
8. Logical Truth	108
VI. Wittgenstein's Fundamental Idea	112
1. Denying the <i>Grundgedanke</i>	113
2. Truth Intrinsic to the Proposition	117
3. The Context Principle	121
4. In Defence of the <i>Grundgedanke</i>	128
VII. Knowledge of the Laws of Truth	135
1. An Understanding of Truth.....	136
2. Truth-Possibilities and Logical Space	139
3. Alice's Thesis and Logical Knowledge	143
4. Two Kinds of Knowledge Revisited.....	146
5. <i>Anamnesis</i>	149
Conclusion	154

I

Introduction: A Priori Knowledge

1. The *Meno* and the Knowledge Within

There is a famous exchange between Socrates and a slave boy in Plato's *Meno*. Socrates draws a square with a side of two feet and asks the boy what the side of another square must be for it to have double the area of the first square. The boy, it seems, does not initially know. Turning to Meno, Socrates emphasizes in what way he will try to help the boy get to the answer.

Look then how he will come out of his perplexity while searching along with me. I shall do nothing more than ask questions and not teach him. (*Meno*, 84c–d)

After a couple of unsuccessful attempts, which the boy himself recognizes as unsuccessful, he realizes that the side of the second square must be the diagonal of the first. The boy did not simply guess the answer correctly but, so to speak, figured it out. What he arrived at is, in one good sense of the word, *knowledge*. The process is familiar enough. The lesson Plato draws from it is less so.

Socrates asks some questions and draws some diagrams in the dirt, these being but prompts to help the boy come to the answer. They are not sufficient to *give* the boy knowledge that a diagonal of a square produces a square with twice the area in the way that looking out the window and noticing water coming down from the sky is sufficient to give one the knowledge that it is raining. Moreover, the boy came to the knowledge not by being corrected by Socrates but by himself recognizing his mistakes as mistakes. Without *already* possessing some geometrical knowledge, this could not be. What coming to an answer in the manner of the slave boy amounts to, Plato suggests, is “finding knowledge within oneself” (85d).

Our focus in this thesis is not on geometrical knowledge but on logical knowledge. But an example parallel to Plato's concerning logical knowledge is familiar to anyone who has taught a logic course. In the first lecture, to give students a sense of what the course will be about, some simple examples of valid and invalid arguments are given. The first is usually something like the following.

If Colonel Mustard was seen near the crime scene, then he is the culprit.

Colonel Mustard was seen near the crime scene.

So, Colonel Mustard is the culprit.

It is generally recognized as valid without a need for explanation. But then a different example is considered.

Either Colonel Mustard is the culprit, or Miss Scarlet is.

Miss Scarlet is not the culprit.

So, Colonel Mustard is the culprit.

Initial intuitive verdicts tend to be mixed about this one. So, the lecturer asks: what does it take for the premises to be true? A student answers. The lecturer continues: so, in supposing the premises true, we are supposing that either the *first* disjunct is true, or the *second* is, and also that the second is *not* true; so, what does that say about the first disjunct? And the student says: “Now I see.” But what do they now see that they previously did not? The explanation of why the argument is valid took the very same shape as the argument itself.

Either the first disjunct is true, or the second is.

The second disjunct is not true.

So, the first disjunct is true.

If someone does not yet know that the first argument is valid, how could the second help them learn it? The second argument can at most serve as a prompt for the students to help them notice what they already have a capacity to notice. If they did not already have the capacity, the explanation could not help them. One can thus teach a logic course by asking questions only. It is puzzling how this could be.

Plato’s answer, if one can attribute the views presented in the dialogues to Plato, is that there are things we never learn but always, in a sense, know. To “learn” truths of geometry or logic is, according to Plato, just to realize that one already knows them. And, as Socrates puts it, “is not finding knowledge within oneself recollection?” (*Meno*, 85d) Plato’s view, in other words, is that our recognition of truths of geometry, mathematics, and logic, is the result of a kind of remembering, *anamnesis*. Later philosophers, including those of today, have defended other accounts. Our goal in this thesis is to defend an account of this kind concerning knowledge of logical truths. Knowledge that it either is or is not raining is an example of this kind of knowledge. Knowledge that any disjunctive syllogism is valid is another.

2. A Priori and A Posteriori Knowledge

On the face of it, Plato seems to think *all* knowledge is ultimately acquired by recollection, that nothing prevents one “after recalling one thing only—a process men call learning—discovering everything else” (*Meno*, 81d), that “searching and learning are, as a whole, recollection” (81d). But a charitable reading is that Plato simply does not apply the term ‘knowledge’ (*episteme*) as broadly as we do in epistemology today. Surely, the slave boy did not acquire the knowledge that Socrates was asking him questions by recollection but by seeing and listening, that is, by experience. We have then a distinction between two kinds of knowledge: knowledge by recollection and knowledge by experience.

The distinction, though it can be found in Plato’s works as well as, for example, in the works of Descartes and Leibniz, became rooted in the mainstream with Kant’s *Critique of Pure Reason*. The question that the *Critique* aims to answer, as Kant states it, is

whether there exists a knowledge that is independent of experience and even of all impressions of the senses. Such **knowledge** is called **a priori**, and it is distinguished from **empirical** knowledge, which has its sources *a posteriori*, that is, in experience. (*CPR*, B2)

Kant further clarifies how he intends the terms ‘a priori’ and ‘a posteriori’ to be understood.

[W]e shall understand by *a priori* knowledge, not knowledge independent of this or that experience, but knowledge **absolutely** independent of all experience. Opposed to it is empirical knowledge, which is possible only *a posteriori*, that is, through experience. (*CPR*, B2–3)

A posteriori or *empirical* knowledge, then, is knowledge that depends on experience. *A priori* knowledge is knowledge that does not. This, broadly speaking, is the characterization adopted in contemporary epistemology and commonly referred to as the a priori-a posteriori distinction.

There is more to Kant’s conception of the distinction, however, than is commonly adopted in contemporary research. Kant’s aim is to assess the tenability of the idea of a priori knowledge that he inherits from an existing tradition. We will cover some of the main tenet of this tradition concerning the a priori-a posteriori distinction which are reflected in Kant’s initial characterization of it. We will not get into details of Kant’s own account.

Kant takes the terms ‘a priori’ and ‘a posteriori’, it seems, from Leibniz who uses them in roughly (so, not exactly) the same way.¹ Taken from Latin, they translate as ‘from what comes before’ and ‘from what comes after’, respectively, which Kant takes to indicate a logical order of priority with respect to experience. A posteriori knowledge, being knowledge of what is experienced, presupposes experience. Whereas experience, in a sense, presupposes a priori knowledge. It is only when the “raw material” of sense impressions, as Kant (B2) describes it, is suitably structured to be capable of supplying us with a posteriori knowledge that it can be construed as experience, properly speaking. It is in that sense that a priori knowledge, according to Kant, must be of what comes “before” experience. Consequently, the possibility of a posteriori knowledge, given Kant’s characterization, depends on the existence of a priori knowledge. Although commonly associated with Kant, we find this idea already in Plato’s *Phaedo*. In an exchange concerning how we might know what it is for one thing to be equal to another, Socrates concludes:

[B]efore we began to see or hear or otherwise perceive, we must have possessed knowledge of the Equal itself if we were about to refer our sense perceptions of equal objects to it, and realized that all of them were eager to be like it, but were inferior. (*Phaedo*, 75b)

Plato maintains, in other words, that for it to be possible to perceive given things as being equal or unequal we must already have knowledge of *what it is* for something to be equal to something else. “We must then possess knowledge of the Equal before that time when we first saw the equal objects” (74e–75a). This point is recognized by commentators. In Plato’s account, as emphasized by Allen (1959), “knowledge of the Forms is epistemically prior to knowledge of particulars” (p. 169).² Closer to Kant’s day, the idea of a posteriori knowledge requiring a priori knowledge is found, for example, in Descartes’ *Meditations*. It comes through clearly in the famous wax example. Noting that all perceivable qualities of a piece of wax change when it is heated, Descartes asks: “So what was there in the wax that was so distinctly grasped? Certainly none of the aspects that I reached by means of the senses. For whatever

¹ Leibniz uses the term ‘a priori’, primarily, to speak of “*a priori* proof” which he characterizes as “proof independent of experience” (“Primary Truths,” in Leibniz 1989, p. 31) as well as more generally of “*a priori* reasons” (“The Monadology,” in Leibniz 1989, §60) which he contrasts with “*a posteriori* reasoning” which is “derived from experience” (ibid., §76). As we will see in chapter III, Leibniz’s characterization of the a priori-a posteriori distinction fits well with Frege’s. For a closer examination of the relationship to Frege’s distinction, see Burge 2000.

² Allen notes the parallel we have drawn between Plato’s and Kant’s accounts, describing Plato’s account as “an infant theory of the *a priori*” (1959, p. 170). Plato’s account is certainly a theory of the a priori. It is not clear why it should be called an infant theory.

came under the senses of taste, smell, sight, touch, or hearing has now changed; and yet the wax remains” (*Meditations*, 30). Thus, Descartes concludes,

even bodies are not, properly speaking, perceived by the senses or by the faculty of imagination, but by the intellect alone, and that they are not perceived through their being touched or seen, but only through their being understood . . . (*Meditations*, 34)

We do, of course, still perceive bodies. Descartes’ point is that doing so presupposes an understanding of *what it is* that is being perceived. We cannot come by this understanding by perceiving bodies, as that would assume that we already can perceive them. It must be *prior* to our perception of bodies, or so Descartes argues.

In Kant’s *Critique*, this logical priority of a priori knowledge is reflected in a characterization of a posteriori knowledge as “a compound” of two aspects.

[I]t is quite possible that even our empirical knowledge is a compound of that which we receive through impressions, and of that which our own faculty of knowledge (sensible impressions merely prompting it to do so) supplies from itself, an addition which we do not distinguish from that raw material until long practice has made us attentive to it and rendered us capable of separating one from the other. (*CPR*, B1–B2)

A posteriori knowledge, Kant proposes (and goes on to argue for), consists, on the one hand, of “raw material” that we “receive through impressions” and, on the other, of “what our faculty of knowledge . . . supplies from itself.” Kant then associates the latter with a priori knowledge. But this calls for some clarification.

We have been using the phrase ‘knowledge of what it is’ to characterize the addition to the “raw material” of sense impressions in a posteriori knowledge. One clearly does, in a sense, need to know what it is for something to be red to be able to know of a given thing that it is red. But this what-it-is knowledge is not yet a priori knowledge. Knowledge of what it is for something to be red may perhaps be thought of merely as a skill or an ability to distinguish red things from things which are not red. A priori knowledge in the sense of ‘knowledge’ that we are interested in is knowledge that can be characterized by a that-clause. The slave boy in *Meno*, Plato argued, already knew *that* a diagonal of one square is the side of another with double the area. The students attending the first lecture in a logic course, it seems, already know *that* $(p \vee q) \wedge \neg q \rightarrow p$.³ Knowledge in this sense is knowledge that implies *truth*. If the what-it-is

³ Following the standard convention, we will use ‘*p*’ and ‘*q*’ as arbitrary declarative sentences. And we will use the standard notation of formal logic to abbreviate logical truths. The abbreviations are the following:

knowledge by itself was sufficient, however, for the sort of knowledge that implies truth, then the resulting truth-implying knowledge would be a candidate for a priori knowledge. This is one way that we might take Kant’s point about learning to separate from the raw material of impressions that which “our faculty of knowledge . . . supplies from itself.” We separate it by recognizing *truths* which are independent of the “raw material.”

Kant’s idea of a priori knowledge being what our faculty of knowledge supplies from itself does not have an obvious origin in Plato’s writing, though it may, perhaps, be read into Leibniz’s work who calls truths we have a priori knowledge of “truths of reason” (*New Essays*, 77) whose “source is in the understanding” (*ibid.*, 75). But it is one thing to maintain, as Plato, Descartes, and Leibniz all do, that reasoning alone—the workings of “our faculty of knowledge”—is sufficient to *uncover* some truths and another to maintain, as Kant seems to, that the *nature* of those truths is in some way explained by the workings of our faculty of knowledge. The latter idea is already a step towards a substantial theory of a priori knowledge.

4. Justification by Experience

One point of caution concerning a priori knowledge, recognized also in contemporary epistemology, is that the absolute independence from all experience that Kant emphasizes must be understood in a specific way. In a priori knowledge, Kant says, “sensible impressions” are “merely prompting” our faculty of knowledge which is thereby “called into activity” (*CPR*, B1). We saw the same idea in play also in Plato’s *Meno*. It is by observing the diagrams that Socrates draws that the slave boy is able to “find the knowledge within himself” (85d). So, a priori knowledge can, in a sense, depend on experience. It may depend on experience in the sense that without experience we might not recognize the truth.

Kant goes further, maintaining that *all* knowledge and thus all a priori knowledge depends on experience in this sense. As he puts it in the opening line of the *Critique*: “There can be no doubt that all our knowledge begins with experience” (*CPR*, B1). There is, Kant maintains, no other way than by experience that our “faculty of knowledge” could be “called

‘ $\neg p$ ’	:	‘it is not the case that p ’
‘ $p \wedge q$ ’	:	‘ p , and q ’
‘ $p \vee q$ ’	:	‘ p , or q ’
‘ $p \rightarrow q$ ’	:	‘if p , then q ’
‘ $p \leftrightarrow q$ ’	:	‘ p if and only if q ’

The connectives are to be taken in their standard truth-functional sense, including ‘if’.

into activity”. It is likely that Kant adopts this assumption from Leibniz, who agrees with Plato that we have a priori knowledge prior to experience but adds that we have it only potentially “as the veins of the marble outline a shape which is in the marble before they are uncovered by the sculptor” (*New Essays*, 86). Experience, in Leibniz’s metaphor, is the sculptor of the mind. Knowledge in the full sense, or “actual knowledge”, Leibniz maintains, is always the upshot of the effects of experience on us.

If there is a sense in which a priori knowledge does depend on experience, we need to clarify the sense in which it does not. That is, we need to specify the role of experience that is characteristic of a posteriori knowledge and distinguish it from its role as a prompt in cases of a priori knowledge. To draw the distinction, we first need to say a little bit about what knowledge in general is.

We have already said that knowledge in the sense that we are interested in implies truth. We might say that, in knowing something, the subject is in contact with the truth. For something to be known in this sense, it must be believed. That is how knowledge gets to be something that a subject has. The knowledge that it is raining is *my* knowledge because I am the one who believes that it is raining. Truth and belief, however, are not sufficient for knowledge. To know something, one must be in contact with the truth *in the right way*. A true belief may be a happy accident. Reading a horoscope is not generally seen as a way to acquire knowledge about the future, even if it sometimes leads to true belief by coincidence. In contemporary epistemology, the third feature of knowledge is known as *justification*. This is where much of our focus concerning the a priori-a posteriori distinction is going to be.

The three conditions—justification, truth, and belief—jointly give us what, since Edmund Gettier’s (1963) influential paper, has been deemed the traditional account of knowledge, commonly referred to as the JTB account. Gettier’s paper is generally agreed to have debunked it. But things are not so simple. An underlying assumption in Gettier’s famous counterexamples as well as in those that were later produced following the same recipe is that we can intuitively recognize justification independently of knowledge. Only then can we set up a case where, intuitively, a subject has a justified true belief but does not know. The goal of an account of knowledge, since Gettier, has been to characterize the intuitive difference between justified true belief and knowledge. But if this has been the goal of an account of knowledge since Gettier, what was its goal before?

In Chisholm's (1957) version of the traditional account of knowledge that Gettier includes among his targets, the justification condition states that the subject must have "adequate evidence" for the truth of what they believe (1957, p. 16). But here we should ask: adequate *for what*? We find the answer in Chisholm's statement of the purpose of his account.

A definition of *knowing that* should be adequate . . . to the distinction between *knowing* and *believing truly*. If I now predict the winner on the basis of what the tea leaves say, then, even though my prediction may be true, I cannot now be said to *know* that it is true. (ibid., pp. 16–17)

Chisholm's account is meant to spell out the respect in which knowledge differs from mere true belief. The subject's evidence that Chisholm appeals to, then, must be adequate for *knowledge*. The strength of the evidence is thus understood *in terms* of knowledge and mere true belief. Likewise, in Ayer's account—Gettier's second target—the justification condition states that the subject must have "the right to be sure" that their belief is true (1956, p. 34). About this right, Ayer says the following:

This right may be earned in various ways; but even if one could give a complete description of them it would be a mistake to try to build it into the definition of knowledge, just as it would be a mistake to try to incorporate our actual standards of goodness into a definition of good. And this being so, it turns out that the questions which philosophers raise about the possibility of knowledge are not all to be settled by discovering what knowledge is. For many of them reappear as questions about the legitimacy of the title to be sure. (ibid.)

So, the right to be sure that Ayer has in mind is to be understood not as something we have an independent pre-theoretic understanding of, but as a placeholder. Given the analogy with the standards of goodness and the definition of good, the placeholder holds the place for the standard of *knowledge*. Like in Chisholm's account, the notion of justification in Ayer's account is thus not an independent component of knowledge alongside truth and belief but simply the *difference* of knowledge and mere true belief, whatever that difference turns out to consist in.

We thus see that the standard account of knowledge as justified true belief involves an understanding of justification that is very different from what is read into it since Gettier. Justification in the standard account is not something we *could* identify independently of knowledge because it is defined as the difference between mere true belief and knowledge. In other words, the unknown in the equation $K = J + T + B$ is not K but J. If we bring the unknown

to the left, as is customary, we get $J = K - (T + B)$. In other words, justification is the difference of knowledge and true belief. This is how we will understand justification in the chapters to follow. Gettier-style counterexamples to the traditional account of knowledge simply cannot get off the ground if things are set up in this way. This does not mean that the cases produced do not show anything interesting about knowledge.

Linda Zagzebski (1994) has argued that Gettier-cases are inescapable. The argument concerns any account of knowledge that does not make the condition of truth redundant by construing the other conditions in a way that they would already imply truth. Zagzebski provides a recipe for debunking any such account. Given the independence of the condition of truth, there is a possible case where the other conditions are met without the belief being true. One can then take this case and add further details to it to make the subject's belief true by coincidence. The result will be a case where, intuitively, the subject does not have knowledge. If correct, then what this means for the traditional view as we now read it is that justification may consist in different things in different cases. As we add details to a case, we might be changing what the difference between knowledge and mere true belief consists in. We thus could not have a single way of spelling out the difference that would be adequate generally. But even if we cannot fully spell it out for all cases, it does not mean that we could not say a great deal about justification.

The contemporary reader may see a parallel between the traditional account of knowledge, or as we understand it now, of justification, and the proposal made more recently by Timothy Williamson (2000) which has come to be known as the *knowledge-first* approach. Williamson gives a general theme of the approach in the opening paragraph of his book.

If I had to summarize this book in two words, they would be: knowledge first. It takes the simple distinction between knowledge and ignorance as a starting point from which to explain other things, not as something itself to be explained. In that sense the book reverses the direction of explanation predominant in the history of epistemology. (2000, p. v)⁴

In line with Williamson's approach, we understand justification in terms of knowledge and so take the notion of knowledge to be the more primitive one in the order of definition. But Williamson goes further, maintaining that knowledge also cannot be characterized as having components. There is a difference between taking knowledge to be definable in terms of other

⁴ As we have now seen, the direction of explanation that Williamson alludes to seems not to have been predominant in the history of epistemology for very long.

things and saying that it cannot be characterized. Given our traditional approach, knowledge *can* be characterized as having components. It is justified true belief. The characterization follows from our definition of justification as the difference of knowledge and mere true belief. And we can investigate the nature of justification as a component of knowledge. This is precisely why the approach will be helpful to us in what follows. By defining justification in terms of knowledge, we have a clear-cut standard of adequacy for a study of justification. Whatever pertains to the difference between knowledge and mere true belief is characteristic of justification, and nothing else is. So, although our approach is like Williamson's in spirit, it is not Williamson's.

Now that we have a definition of justification, we can clarify the sense in which a posteriori knowledge depends on experience but a priori knowledge does not, even though experience may in some sense be involved in both. A posteriori knowledge is knowledge where experience is playing a *justifying* role. When experience is involved in knowledge in some other way, we will say it is playing a *merely enabling* role. Although experience does not play a justifying role in a priori knowledge, it may play a merely enabling role in it. Experiences required for learning the meanings of words or for doing calculations or proofs with pen and paper are of that sort.

Our definition of justification lets us state clearly when experience is playing a justifying role in knowledge and thus when knowledge is a posteriori and when a priori. Since justification, as we have defined it, is the difference of knowledge and mere true belief, a given experience is playing a justifying role in knowledge if and only if it cannot be explained why the subject has knowledge rather than a mere true belief without mentioning the experience. When the explanation can be given without mentioning the experience, then that experience is at most playing a merely enabling role. Consequently, knowledge in a given case is a priori if and only if it can be explained why the subject has knowledge rather than a mere true belief without mentioning any experience. Otherwise, the knowledge is a posteriori.

We should note that nothing about our characterization of the justifying role of experience implies that subjects themselves should be able to tell whether their knowledge depends on experience or not, let alone be able to cite which experience their knowledge depends on. In fact, the experiences that we need to appeal to in the explanation needn't even be the subject's own experiences. If scientists discover that there is life on Saturn's moon Titan, then I too will eventually come to know that there is life on Titan. But I will not observe how

things are on Titan or gather any evidence of life there. I may observe the newspaper article where the discovery is announced. But appealing to those observations may help to explain why I have the *belief* that I do but not why this belief is knowledge. To explain the latter, we would need to talk about scientists having found strong evidence of life on Titan. That is, we would need to appeal to the observations that the *scientists* make.

We now have a relatively clear-cut test for whether knowledge in a given case is a priori or a posteriori. The use of this test is of course cumbersome as it involves going through experiences which are involved in a given case of knowledge one by one. But the test, even if not applied, helps to delineate the distinction between a priori and a posteriori knowledge more precisely. The claim we will go on to defend in the coming chapters is that to explain of any given logical truth why someone has knowledge of it rather than a merely true belief we never have to appeal to experiences of any sort. But just this seems to raise a further question that we have yet to articulate.

If there are no experiences to appeal to in an explanation of why the person is justified in holding a logical belief, what *is* there to appeal to? So far, we have characterized a priori knowledge in negative terms only. A priori knowledge, given our characterization, is simply knowledge that is not a posteriori. A need to supplement this with a positive characterization is generally seen as the most pressing issue for proponents of the a priori-a posteriori distinction. Michael Devitt, a vocal critic of the idea of a priori knowledge, has put the worry clearly.

[T]he whole idea of the *a priori* seems deeply obscure. *What is it* for a belief to be justified *a priori*? What is the nature of this non-empirical method of justification? Without satisfactory answers the *a priori* is left mysterious. (Devitt 2005, p. 106)

What is left mysterious in the absence of a positive characterization of a priori knowledge is why a priori knowledge, despite not depending for its justification on experience, is still knowledge. In other words, given Kant's characterization of the distinction, we need to provide an account of what the justification that is involved in cases of a priori knowledge consists in. This, as we will see in chapter III, is commonly taken as a call to identify something else that could play the same role that experience plays in a posteriori knowledge. We will also see that it is not the only way to respond to the worry.

5. Why A Priori?

What reason is there to think that we have a priori knowledge, knowledge where no experience is playing a justifying role? In contemporary epistemology, it has become standard by proponents of the a priori to see the distinction between a priori and a posteriori knowledge as an intuitive distinction that we have learned to make in our everyday lives as knowers. The short answer then is that the existence of a priori knowledge is simply a datum.

Our knowledge of logical truths is a clear example of knowledge that appears clearly to be independent of experience. As Wittgenstein famously puts it in the *Tractatus*, “I know nothing about the weather when I know that it is either raining or not raining” (*TLP*, 4.461). When considering simple logical truths like ‘it is or is not raining’, we are simply at a loss to cite any experience that might justify our belief in their truth. The state of the weather makes no difference to the truth of ‘it is or is not raining’. The proposition is true when it is raining as well as when it is not.⁵ And so, checking the weather seems to play no role in coming to know that it is true. But what else is there to check? It seems that to learn that it is or is not raining, what we need to do, instead of looking around, is to simply figure it out, so to speak. And what holds of simple logical truths holds of all of them, though the process of figuring out that the proposition is true will in some cases require much more effort. Truth-tables and proof-rules are notational aids which have been developed to make this task easier and less prone to error, much like a telescope has been developed to enhance the powers of the human eye to help us observe distant heavenly bodies. The use of notational aids, of course, does involve experience. But experience is there playing a merely enabling role. The observations of truth-tables on paper no more justify our logical beliefs than observations *of* a telescope justify one’s beliefs about distant heavenly bodies.

A potential problem with working from intuitive examples of a priori knowledge arises from our characterization of the justifying role of experience. Experience is playing a justifying role in knowledge when it cannot be explained why the subject knows rather than has a mere true belief without appeal to experience. This leaves open the possibility that subjects themselves are unaware of whether or how their knowledge depends on experience.

⁵ Here and elsewhere I use the term ‘proposition’ to mean an *expression* of a truth-condition. It is in this sense of ‘proposition’ that truth-functional logic, for example, is sometimes called *propositional* logic. And it corresponds to how the term (or its German equivalent ‘*Satz*’) was used by many of the authors we will be considering throughout this thesis. It differs, however, from the widespread use of the term in contemporary philosophy to mean what a sentence says. In place of the contemporary term, we can use ‘propositional content’ (if we must). I will clarify the notion of a proposition in chapter V.

Consequently, the intuitive examples of a posteriori and a priori knowledge may be deceptive. Though it appears that in some of them, experience is playing a special role while in others it does not, this appearance has no authority on its own. It may turn out that logical knowledge depends on some experience that we simply haven't thought to consider.

One general theme that emerges in the philosophical tradition is that there is also another way of approaching the matter, not by intuitive examples but by argument. We saw that Plato and Descartes both maintained that there are things that we evidently know but *cannot* know by experience. To argue for this is to argue that some of our knowledge is independent of experience regardless of whether it appears to be, or, as it is with Descartes's argument concerning our knowledge of the persistence of bodies, *despite* the appearance that our knowledge does depend on experience. We have seen that one way this might be defended, as is reflected in Kant's characterization of a posteriori knowledge, is to argue that the possibility of any a posteriori knowledge presupposes a priori knowledge.

For illustration, we may sketch an argument in this spirit for our simple case of logical knowledge. Knowledge that it is raining is a posteriori. The state of the weather must be checked—and so, observations must be made—to establish that it is raining. This we evidently can do. But to check whether it is raining is to determine which of the two alternatives—that it is raining or that it is not raining—is actual. But this means that to even take up the task of checking whether it is raining, we should in some sense already know that those are the only two alternatives, and that they cannot both be actual. That is, we must in some sense *already* know that it is or is not raining, and that it is not both raining and not raining. We must, in other words, already have some logical knowledge. Only if we know what possibilities there are *could* we proceed to establish which of them is actual. We can construct a similar argument for any knowledge by experience. But, if having knowledge of some logical truths is required for gaining any knowledge by experience, then knowledge of logical truths cannot itself depend on experience. It must be a priori, regardless of whether it appears to be.

The argument is of course not meant to be decisive or even persuasive (yet) but is meant only to illustrate the sort of approach that is possible to take to explain why we have a priori knowledge. But the argument also illustrates a fundamental limitation of this kind of approach. Logic is by no means the only area where knowledge has commonly been thought to be a priori. Mathematical knowledge is another candidate, as is moral knowledge and philosophical knowledge. Moreover, the range of the a priori is not limited to what is commonly been

understood as such. We saw Descartes argue that our knowledge of the existence of bodies is a priori. Likewise, Plato is known for the idea that to know aesthetic truths it is necessary and sufficient to comprehend the form of Beauty—“that it is through Beauty that beautiful things are made beautiful” (*Phaedo*, 100e)⁶—which, given his general framework we have briefly discussed, would make all aesthetic knowledge a priori. It would be remarkable if there was a single (and sound) argument that decided the truth or falsity of each of these proposals. The approach by argument is thus bound to have only a limited scope. It is for this reason, that we have set out to argue for the apriority of logical knowledge only.

6. The Aim and Structure of the Thesis

In the chapters to come, I will argue that knowledge of logical truths is a priori knowledge. The strategy I will pursue follows the second of the two approaches—the approach by argument—outlined in the previous section. I will argue that the possibility of recognizing something as truth-evaluable presupposes that one already has knowledge of every logical truth. In other words, once we as theorists understand truth properly, any apparent difference between having logical knowledge and having an understanding of truth will vanish. Since experience is a way of recognizing what is true, logical knowledge cannot itself depend on any experience and must therefore be a priori. This account will need no positive characterization of what justification in logical knowledge consists in. One’s logical beliefs will always come out vacuously justified whenever they are true. There is, I will argue, no difference in logic between knowledge and true belief.

The aim of this thesis may come across as odd to a contemporary reader in several respects. One is due to widespread reservations in contemporary epistemology concerning the very idea of a priori knowledge. I will address the most common objection in this spirit, an objection based on Quine’s (1951) work, in chapter II and argue that the objection is not as strong as it is often taken to be. This opens the way at least for the idea that some of our knowledge *might* be a priori. But even those already sympathetic to the possibility might find the goal of the thesis puzzling. The a priori-a posteriori distinction is normally construed as a distinction between different ways of having or gaining knowledge. Thus understood, the

⁶ Here, Plato means that it is through Beauty *alone* that beautiful things are beautiful. As Socrates says in the same passage, “I no longer understand or recognize those other sophisticated causes, and if someone tells me that a thing is beautiful because it has a bright color or shape or any such thing, I ignore these other reasons” (100d).

distinction is primarily a cognitive distinction and may have something to do with what is known only in a derivative sense, if at all. For example, the same truth, if it is commonly maintained, can be known both a priori and a posteriori. The claim that all logical knowledge is a priori is in tension with that. In chapter III, I will take up this issue and argue that to construe the a priori-a posteriori distinction as a distinction between ways of knowing is a mistake, that to have a distinction between a priori and a posteriori knowledge that is of epistemic significance we must construe it as concerning kinds of truth. This further clarifies why our focus must be limited to knowledge concerning a specific range of truths. But the ideas presented in chapter III are hopefully also of some independent interest as they challenge the conception of apriority which has been standard since Saul Kripke's (1980) influential work.

In chapter IV, I will sketch an outline of how features of logical truths could lead to an account of the apriority of logical knowledge. The inspiration for the idea comes from a view famously held by Alice in Carroll's *Through the Looking-Glass*, that one cannot believe impossible things. Restricting this to logical impossibility, I will clarify Alice's claim by building on similar ideas endorsed by Berkeley, Kant, and others. I will trace the source of Alice's view to the idea that the truth-values of propositions of logic, of logical truths and falsities, do not depend on how things are. As we will see, it is a consequence of this that logical falsities—propositions whose truth is logically impossible—cannot be believed, that there is nothing in logical falsities *to* believe. Moreover, knowledge of a logical truth would consequently be knowledge of a special kind. It would be knowledge that could not be acquired by checking how things are, whether by experience or otherwise. It would, in other words, be a priori knowledge.

With an outline for a project specified, I will proceed to fill it in. In chapter V, I will provide an account of logical truths as determined by the nature of truth. The idea goes back to Frege (and perhaps further back to Kant) but is also commonly adopted in logic textbooks today where logical consequence is defined with the help of a Tarski-style definition of truth. I will argue that the standard Tarski-style definition has problems and will adjust the definition accordingly. The new definition will be developed following Frege's ideas more closely. The resulting definition of truth will not make use of a distinction between satisfaction and truth as Tarski's definition does. It will also define the truth of atomic propositions differently from how it is done in most contemporary logic textbooks (though not in Tarski's own work). Building on Frege's account of generality and the work of Wittgenstein and Ramsey, I will construe an atomic proposition as a concatenation of terms each of which contributes an object

to a truth-condition. This ties the nature of all truth to reference in a way that the standard account of atomic propositions does not.

In chapter VI, I will assume the definition of truth developed in the previous chapter to defend, at least in outline, Wittgenstein's account of logical truths as it is presented in the *Tractatus*. The account is commonly summarized by the dictum that logical truths are "empty," or *sinnlos*, as Wittgenstein describes them. If correct, it implies that the truth-values of propositions of logic are not answerable to how things are but are instead intrinsic to those propositions. The defence of the account comes down to a defence of what Wittgenstein calls his fundamental idea, or *Grundgedanke*, that signs for the logical operations cannot be construed as terms and thus cannot be contributing components to truth-conditions. To defend the *Grundgedanke*, I will first defend a specific reading, endorsed by Anscombe, of Frege's famous Context Principle. In short, I will argue that any given term can be assigned a referent only when considered in combination with other terms that jointly make up a sentence. From this and the definition of truth developed in the previous chapter, I will then prove the *Grundgedanke* and thereby the account of logical truths endorsed by Wittgenstein.

In chapter VII, the final chapter of the thesis, I will return to the outline sketched in chapter IV and retrace the steps laid out in it. Taking inspiration from Dummett's account of generalizations, I will first explain what is involved in an understanding of what it is for anything to be true—that is, of what truth is—and then proceed to show that this understanding is both necessary and sufficient for the possession of logical knowledge. First, I will specify the broadest range of possibilities there is, which includes everything there is to believe. Given the *Grundgedanke*, logical falsities do not fall within that range and are therefore not candidates for belief. This is the argument for Alice's thesis. I will then proceed to develop an account of logical knowledge based on this result. In short, it will become clear that to have knowledge of logical truths just is to have a grasp of the broadest range of possibilities which in turn is tantamount to having an understanding of what truth is. The ability to recognize any proposition as true—since it presupposes an understanding of truth—therefore also presupposes knowledge of every logical truth. One therefore cannot gain knowledge by experience unless one already has logical knowledge. One question left open in chapter IV is what justification in logical knowledge consists in. This question will be answered in chapter VII. Justification in logical knowledge will turn out to be as vacuous as the truth of logical truths.

The last section of chapter VII addresses an objection that looms over the chapter and indeed over much of the second half of the thesis. It is the third respect in which the aim of this thesis may strike the reader as odd. It will become clear that logical knowledge cannot be acquired but must always already be possessed. This does not seem to fit our normal understanding of logical knowledge as something that one can fail to have and struggle to acquire. To account for the appearance, I will introduce a second notion of knowledge which, unlike the normal notion, concerns the truth of the proposition by means of which the knowledge is characterized and is thus partly linguistic. Knowledge in this partly linguistic sense will inherit its epistemic status from the underlying regular knowledge. The appearance that logical truths can be learned is then accounted for as the possibility of finding that a given proposition expresses the knowledge one already has. The eventual account will closely resemble Plato's account of knowledge of Forms in terms of *anamnesis* that we briefly covered above.

II

Logic and the Web of Belief

We have characterized the a priori-a posteriori distinction as it has been construed historically. What has remained of this characterization in contemporary epistemology is that a posteriori knowledge depends on experience and a priori knowledge does not. But even under this minimal characterization, the distinction has lost much of its popularity, for there is a common objection to the idea of a priori knowledge as such, regardless of any specific account of it. The objection comes from Quine's critique of a different distinction, the analytic-synthetic distinction, presented in "Two Dogmas of Empiricism" (1951) and elaborated more recently by Michael Devitt.

In this chapter, we will consider the Quinean objection to apriority and see why it is not as strong as it is often thought to be. We will first take a closer look at how the objection arises in connection to analyticity and what reason there might be, as Putnam (1983) has suggested there is, to extend Quine's points about analyticity to apriority. As we will see, there is none. In short, for Quine's objection to go through, a priori knowledge must already be construed as a peculiar kind of a posteriori knowledge, knowledge where any experience can play a justifying role. Finally, we will consider Devitt's (2005) more recent version of the Quinean argument, that the positive account of logical knowledge that Quine offers, after some adjustments are made to it, is (at least so far) the best explanation of how our logical beliefs get to be knowledge. We will see that it is not. Seeing why the Quinean objection fails opens the way for us to explore accounts of a priori knowledge in the next chapter.

1. Quine's Confirmational Holism

Like the distinction between a priori and a posteriori knowledge, the analytic-synthetic distinction was popularized by Kant in *Critique of Pure Reason*.

In all judgments in which the relation between subject and predicate is thought . . . this relation is possible in two ways. Either the predicate B belongs to the subject A as something which is (covertly) contained in the concept A; or B lies outside the concept A, though connected with it. In the former case I call the judgment **analytic**, in the latter **synthetic**. (*CPR*, A6)

This characterization has been found, at least since Quine, to call for clarification. In particular, it is unclear how to understand the idea of the subject being contained in the concept. Kant's own characterization of the analytic-synthetic distinction has consequently not become widespread. The distinction is instead primarily seen as an intuitive distinction between two kinds of propositions. There is an intuitive difference in how the truth-values of propositions like

‘All bachelors are unmarried’,

‘All bachelors are untidy’

are determined. The truth of the latter concerns the habits of each individual bachelor whereas the truth of the former has something to do with the special way in which bachelorhood is related to marital status. An account of analyticity is supposed to clarify what this special relationship consists in. In “Two Dogmas”, Quine sets out to show that all such accounts given so far are unsatisfactory. Quine has two targets in particular. One is an account which he attributes to Kant but is more appropriately traced back to Frege. The other account Quine attributes to the logical positivists. Our focus will be on Quine's objection to the second target but before we turn to it, we must mention the first one briefly.

Given Kant's characterization of the analytic-synthetic distinction, the special way in which bachelorhood is related to marital status is that the predicate *being unmarried* belongs to bachelorhood by being covertly contained in the concept of bachelorhood. Quine proposes to spell out this proposal in terms of linguistic meaning. What Kant's proposal then amounts to is the idea that ‘all bachelors are unmarried’ is true because ‘bachelor’ means the same as ‘unmarried man’. A proposition is analytically true, given this account, if and only if substituting synonyms for synonyms in it results in a logical truth. ‘All bachelors are unmarried’ is analytic by this characterization because ‘bachelor’ is synonymous with ‘unmarried man’ and the substitution of one for the other in the proposition would yield ‘all unmarried men are unmarried’ which is a logical truth.

The idea closely resembles Frege's account of analyticity in *Foundations of Arithmetic* (§3) where Frege takes an analytic proposition to be a proposition whose truth follows solely from basic axioms and definitions. Although Kant's characterization might be motivated partly by propositions like ‘all bachelors are unmarried’, Frege's certainly was not. Frege's aim was to show that all mathematical truths are derivable from logical axioms. But, since definitions are common in mathematics, the derivations will not be possible unless one appeals to

definitions in the derivation. What sets Frege's account apart then is that the synonymy relations are always stipulated in terms of explicit definitions. As Quine (1951, p. 26) himself recognizes, there is no mystery about how this works. The definitions simply serve as further propositions alongside the axioms from which to derive new propositions.

Quine adds, however, that a *general* account of analyticity given in terms of explicit definitions is inadequate since the truth of many (apparently) analytic propositions, including 'all bachelors are unmarried', is not known by inference from definitions.⁷ Quine then goes on to show by way of examples that there also does not seem to be any other straightforward way of explaining synonymy which does not implicitly appeal either to synonymy itself or to analyticity. Both would make the characterization of analyticity in terms of synonymy circular. We will not be assessing the strength of this criticism here.⁸ Where the story becomes relevant to us is when Quine, having covered all the straightforward ways of explaining synonymy, turns to something less straightforward, namely to the verificationist theory of meaning held by the logical positivists and which, arguably, was Quine's real target all along.

The verificationist theory of meaning, as sketched by Carnap (1932), follows a pattern familiar from truth-conditional semantics (that we will cover in more detail in chapter V). The meanings of all sentences are assumed to be determined compositionally by the meanings of some basic sentences. But rather than taking the meanings of basic sentences to be given by conditions under which they are true, verificationists posit that it is given by conditions under which the basic sentences are *verified*.⁹ The verification-conditions of basic sentences are in turn taken to bestow meaning onto the expressions that make up the basic sentences. This would ensure that all non-basic sentences in which the expressions occur also have meanings. All meaning is thereby understood in terms of verification-conditions of basic sentences (called protocol sentences or observation sentences).

⁷ For example, 'bachelors' does not really seem to mean the same as 'unmarried man'. A bit of reflection on the matter reveals that 'bachelor' does not seem to apply to men who are not yet of marriageable age. So, an age-range seems to be built into the meaning of 'bachelor'. Reflecting on the matter further, we find that the Pope does not seem to be a bachelor, either. What about widows? The more examples we consider the less likely it seems that we have anything so definite in mind which could be formulated as an explicit definition of 'bachelor'.

⁸ Although see Grice & Strawson (1956) for a response.

⁹ Verificationists did recognize truth-conditions but did not see a significant difference between truth- and verification-conditions, holding either that the two notions effectively amount to the same (Carnap 1932, p. 62) or that talk of truth-conditions is itself to be understood in terms of verification (Schlick 1932/33, p. 87).

One promise of verificationism was that it could help to provide a precise characterization of the analytic-synthetic distinction. The eventual goal, as Hempel (1950) puts it, was to show that

[a] sentence makes a cognitively meaningful assertion, and thus can be said to be either true or false, only if it is either (1) analytic or self-contradictory or (2) capable, at least in principle, of experiential test. (p. 108)

Verificationism was meant to explain the second category of propositions which was identified with the category of synthetic propositions. Every proposition that could not be shown to be either analytic or synthetic was deemed nonsensical. There was considerable disagreement among the logical positivists over how best to understand the basic sentences and the verification-conditions assigned to them.¹⁰ As we will see, Quine in his criticism assumes one conception that was on the table at the time, the conception maintained by Moritz Schlick. “The *meaning* of every proposition,” Schlick says, “is finally to be determined by the given, and by nothing else” (1932/33, p. 87), where, by ‘the given’, Schlick means “the occurrence of certain sense-impressions” (ibid., p. 90). Given Schlick’s conception, verification-conditions may be thought of as sets of verifying experiences—experiences the occurrence of which by itself verifies the proposition. Since the absence of a verifying experience (when one is in deep sleep, for example) is clearly not enough to falsify a proposition, we should also recognize a set of falsifying experiences for each proposition. Each basic proposition is thus construed as having one associated set of possible experiences which verify the proposition and another set of experiences that falsify it. The two sets are then taken to jointly determine the meaning of the basic proposition.

There are two ways of connecting verificationism to the analytic-synthetic distinction. Verificationism, since it is a theory of meaning, might explain synonymy—sameness of meaning—in a non-circular way. But Quine’s primary reason for focusing on verificationism is that he takes an account of analyticity to be derivable from verificationism directly without a detour through the Frege-style characterization. Analytic propositions, Quine suggest, could be treated as limit cases where all possible experiences either verify the proposition or all possible experiences falsify it. By the time Quine was writing “Two Dogmas” even the proponents of verificationism no longer held it. But, Quine notes, even when verificationism is rejected “the notion lingers” that each proposition has its characteristic sets of possible

¹⁰ For an overview of the disagreement see Davidson 1982.

experiences which speak for or against its truth (1951, p. 38). This lingering notion alone is sufficient to draw the analytic-synthetic distinction. As Quine puts it:

[A]s long as it is taken to be significant in general to speak of the confirmation and infirmation of a statement, it seems significant to speak also of a limiting kind of statement which is vacuously confirmed, *ipso facto*, come what may . . . (ibid.)

By a “kind of statement which is vacuously confirmed . . . come what may”, Quine means analytic truths as they are understood in light of the lingering notion. If each proposition has its associated set of possible experiences which verify it, then there may as well be propositions where the set is maximal. There is no need to take the meaning of the proposition to be given by that set. Of course, the analytic-synthetic distinction would then not be associated with meaning as Quine initially assumed, but if the distinction is primarily an intuitive distinction, then it needn’t be construed in terms of meaning.

Quine’s objection to the lingering notion is that it is blatantly false. There are no characteristic sets of verifying experiences associated with propositions. When confronted with recalcitrant experience—experience which does not fit our expectations based on what we believe—we have a *choice* to make. There are many propositions we believe to be true, and there are consequently many ways of relieving the tension between recalcitrant experience and what we believe. A belief is disconfirmed by experience when it is given up in the face of it so that the tension is relieved. But it is not with an individual belief that experience is then in tension but with the totality of our beliefs—“our statements about the external world,” as Quine puts it, “face the tribunal of sense experience not individually but only as a corporate body” (1951, p. 38). The view suggested by the rejection of the lingering notion is *confirmational holism*, summarized by Quine in the following famous passage in the final part of “Two Dogmas”:

The totality of our so-called knowledge or beliefs, from the most casual matters of geography and history to the profoundest laws of atomic physics or even of pure mathematics and logic, is a man-made fabric which impinges on experience only along the edges. Or, to change the figure, total science is like a field of force whose boundary conditions are experience. A conflict with experience at the periphery occasions readjustments in the interior of the field. Truth values have to be redistributed over some of our statements. Re-evaluation of some statements entails re-evaluation of others, because of their logical interconnections—the logical laws being in turn simply certain further statements of the system, certain further elements of the field. (p. 39)

Confirmation and disconfirmation, given the holistic view, does not primarily concern the truth of a given proposition but of finding an *equilibrium* between incoming experience and how we distribute truth-values over all propositions. Experience that does not disrupt the equilibrium is confirming experience, and experience which does is disconfirming. But the experience itself nor any one of the propositions which we believe to be true settles *what* the experience disconfirms. All that is rationally required of us in the face of recalcitrant experience is the restoration of the equilibrium. But no one way of restoring it is forced. And though some ways may be more difficult to carry out, requiring more drastic changes in the field, Quine maintains that all ways of restoring the equilibrium are equally rational, since, on the holistic view, maintaining an equilibrium is all there is to rationality. In contrast, the lingering notion implies that rationality consists in giving up specific beliefs in the face of specific experiences: one synthetic proposition is confirmed or disconfirmed by some specific experiences and another by others. Given holism,

it becomes folly to seek a boundary between synthetic statements, which hold contingently on experience, and analytic statements, which hold come what may. Any statement can be held true come what may, if we make drastic enough adjustments elsewhere in the system. (1951, p. 40)

It is Quine's confirmational holism, developed in response to the lingering notion, that later philosophers like Putnam (1983) and more recently Devitt (2005) came to see as providing for a general objection to the existence of a priori knowledge.

2. A Problem for A Priori Knowledge?

A connection between Quine's confirmational holism and the a priori-a posteriori distinction was first proposed by Putnam (1983).

Confirmation, in the positivist sense, has something to do with rational belief. A statement which is highly confirmed is a statement which it is rational to believe, or rational to believe to a high degree. If there are indeed statements which have the maximum degree of confirmation in all circumstances, then these are simply truths which it is *always rational to believe*, nay, more, truths which it is never rational to even begin to doubt. Many philosophers have believed that there are such truths. Perhaps this is what Aristotle thought a *first principle* was like; more likely it is what Descartes thought a *clear and*

distinct idea was like. On the face of it, then, the concept of a truth which is confirmed no matter what is not a concept of *analyticity* but a concept of *apriority*. (p. 90)

Putnam rightly notes that confirmation (that is, verification) in the logical positivist sense has to do with the rationality of belief. In our discussion, we have not elaborated on what rationality is nor given it an explicit role in knowledge. For Putnam's point to bear on apriority, we should switch to talk of justification—the difference of knowledge and mere true belief. (We may suppose questions of rationality to bear on this difference.) What Putnam emphasizes is that verificationism, though explicitly stated as a theory of meaning, doubles as an epistemological theory. It characterizes the meanings of basic propositions in a way that makes them fit a specific account of how their truth is known. The lingering notion that Quine criticizes is the epistemological remainder of verificationism, once the aim to explain meaning is given up. Quine is of course aware of this, as he advertises his alternative—confirmational holism—as an epistemological theory, “empiricism without the dogmas” which concerns “[t]he totality of our so-called knowledge” (1951, p. 39).

Putnam however is wrong to suggest that the target of Quine's critique—the idea that some propositions, by their very nature, are confirmed no matter what—is an account of apriority. The lingering notion that Quine's holism is meant to be an alternative to is that there is a characteristic set of *verifying experiences* associated with a proposition, and for some propositions the set is maximal, that is, they are propositions verified by all experiences. That is the sense in which the lingering notion implies that some propositions may be held come what may. Crucially, these propositions are not construed as not calling for verification by experience. They may be rationally held in the face of any experience precisely because *every* experience verifies them. This is not a priori knowledge by any measure. Nor was a view like this held by logical positivists about analytic propositions. It would imply that all analytic truths are known in one go and, given that verificationism is a theory of meaning, it would also imply that they all have the same meaning. A proper assessment of the logical positivist approach is given by Dummett.

On the positivist view, logically necessary statements do not have an ordinary sense at all: for no sense experience could falsify them, and hence no sense-experience is needed in order to verify them; if they are to be conceded a sense, this must be a sense of an entirely different kind from that possessed by empirical statements, a sense not to be explained in the same way, in terms of the notions of verification and falsification. (*Pol*, p. 588)

In fairness to Putnam, the mistake is already there in “Two Dogmas”. It originates from Quine’s attempt to construe analyticity in terms of verification alone. To quote the passage again,

as long as it is taken to be significant in general to speak of the confirmation and infirmation of a statement, it seems significant to speak also of a limiting kind of statement which is vacuously confirmed, *ipso facto*, come what may (p. 38)

Logical positivists never assigned any epistemological significance to this sort of vacuous confirmation. Although verificationism may be thought to *delineate* the range of analytic truths, it cannot explain what knowledge of those truths consists in. So, Quine’s explicit target is a strawman. But there is another reason for thinking that Quine’s confirmational holism poses a challenge to the idea of a priori knowledge, even if the idea was not his original target.

We first need to be clear about the sense in which confirmational holism, as Quine presents it, is holistic. It is the view that our beliefs “face the tribunal of sense experience *not* individually but *only* as a corporate body”—that “[t]he unit of empirical significance is the whole of science” and not any proper part of it (1951, p. 166, emphasis added). To read confirmational holism as a theory of knowledge, we should understand the idea of empirical significance and “the tribunal of sense experience” as concerning justification by experience. Quine’s claim then is that justification by experience is not the sort of thing that concerns individual beliefs. It concerns only the totality of all our beliefs. Thus understood, the immediate target of holism is not a priori knowledge, as Putnam suggested, but *a posteriori* knowledge. In place of individual beliefs being justified by experience, what we are offered instead is a “thorough pragmatism” (p. 43).

Each man is given a scientific heritage plus a continuing barrage of sensory stimulations; and the considerations which guide him in warping his scientific heritage to fit his continuing sensory promptings are, where rational, pragmatic. (ibid.)

The dialectical reason why Quine’s critique concerns a posteriori knowledge and not a priori knowledge is that his target in “Two Dogmas” is the lingering notion left by verificationism. That notion was an *account* of a posteriori knowledge. But because of this, it is also an overstatement to say that confirmational holism is a rejection of the very idea of justification of individual beliefs by experience. It is a rejection only of a specific elaboration of it where the occurrence of any given experience itself selects for beliefs which are thereby verified or falsified. If this account was right, then a subject could only be a passive receiver of knowledge. Once an observation is made—once the subject has opened their eyes, so to speak—the

propositions characterized by the resulting experiences would automatically be known by the subject. Quine rightly rejects this picture, noting that a subject is not a passive receiver but an active pursuer of knowledge who weighs evidence, considers alternative hypotheses, questions experiments, etc. Once we make room for this role of the subject, confirmational holism is the result. Confirmational holism is thus better understood not as a rejection but as an alternative account of a posteriori knowledge with a more sophisticated characterization of what justification by experience consists in.¹¹ But this raises an exegetical question.

If the target of Quine's attack is an account of a posteriori knowledge, then where do Quine's claims about the laws of logic being "simply certain further statements of the system" (1951, p. 39) come from? Given the lingering notion that Quine criticizes, all experiences vacuously verify logical truths. Moreover, as Quine mistakenly takes the idea, this vacuous verification by experience is what accounts for the justification of our logical beliefs. So, Quine *starts out* with the assumption that logical truths can be verified by experience. The aposteriority of logic is common ground between the strawman that Quine attacks and the view he arrives at. When Quine recognizes the role of the believing subject in verification by experience, the vacuous verification from before is no longer vacuous. When subjects were previously construed as being, as it were, saddled with justification to believe logical truths in the face of any experience, they are now construed as having a say in whether to retain a logical truth in the face of experience or not. So, logical truths do not receive any special status in the web of belief because knowledge of them is treated by Quine as a special case of a posteriori knowledge from the start. When the verificationist account of a posteriori knowledge is replaced with a subject-centred holistic account, knowledge of logical truths is no longer seen as a special case, but it *remains* seen as a posteriori. Thus, not only does Quine's holism provide us with no reason to suppose that logical truths might be a posteriori, but it also provides no reason for the logical positivists to suppose that it challenges their view of analytic truths. A lot has been read into Quine's claims about logic in "Two Dogmas" that simply isn't there.¹²

¹¹ For a detailed formulation of confirmational holism as an account of a posteriori knowledge, see Quine & Ullian, *Web of Belief* (1978). The resulting account is much closer to verificationism than "Two Dogmas" might suggest. Quine and Ullian recognize the need for observation sentences much like in verificationism. But, since the semantic aim has been given up, the role of observation sentences is no longer to give meaning to the basic expressions of the language. Their role is to convey the initial empirical data to be explained by hypotheses. On this view, the metaphor of the periphery and the interior of the web is spelled out by the recognition of two sorts of a posteriori knowledge. There is knowledge of the data itself, that is, of the truth of the observation sentences, and there are bodies of knowledge where the beliefs jointly help to explain the data better and make better predictions (thereby giving us new observational knowledge) than the alternatives.

¹² In any case, Quine (1986) seems later to distance himself from the radical form of holism, maintaining that the possibility of translation requires that basic logic be assumed to remain constant across languages.

3. Empiricism by Abduction?

Alas, Quine is now often seen as having shown all knowledge, logical knowledge included, to be a posteriori. Since we have no argument from Quine to this effect, we need to turn elsewhere to fill the void. We find an argument in a contemporary defence of holism by Michael Devitt (2005). Building on Quine's remarks about logic in "Two Dogmas", Devitt takes confirmational holism to be providing us with "an alternative naturalistic account of the troublesome examples of allegedly *a priori* knowledge" (2005, p. 105), under which Devitt classifies cases of logical knowledge. By setting as his target "troublesome examples," Devitt assumes that the main reason for supposing logical knowledge to be a priori is that it is hard to see how examples of logical knowledge could involve dependence on experience. Against this reason, Devitt presents an argument by abduction.

[A] naturalistic critic of the *a priori* faces two tasks: to undermine the motivation by showing that the troublesome knowledge could be empirical after all; and to demonstrate the obscurity of the *a priori*. Success in the second task would show that an *a priori* explanation of the troublesome knowledge, indeed of anything, was very unpromising. Success in the first task would show that an empirical explanation was available. So we would have a nice abduction for naturalism: the *best* explanation of that knowledge is that it is empirical. (2005, p. 106)

The absence of any argument in Quine's "Two Dogmas" against the apriority of logic is thus no obstacle for Devitt's project. If confirmational holism *can* incorporate logical knowledge as a posteriori knowledge, then all that is needed for it to be the best account of logical knowledge is for it to be better than the alternatives which construe logical knowledge as a separate kind of knowledge. In fact, demonstration of the obscurity of a priori knowledge may not even be necessary. Considerations of simplicity would already speak in favour of all-encompassing empiricism. As Devitt himself puts it:

It is overwhelmingly plausible that *some* knowledge is empirical, "justified by experience." The attractive thesis of naturalism is that *all* knowledge is; there is only one way of knowing. (p. 105)

The problem for this approach, as we have seen, is that not only does Quine not give us an argument for the aposteriority of logical knowledge, but he also gives us no account of how logical knowledge *could* be a posteriori. He says that it is, but we have not been told what to make of those claims, apart from one example:

Revision even of the logical law of the excluded middle has been proposed as a means of simplifying quantum mechanics (Quine 1951, p. 40)

What Quine suggests is that rejection of a logical law to restore equilibrium with incoming experience is to be understood as an adoption of a deviant logic to simplify an empirical theory.¹³ One well-known proposal in this spirit has been made by Birkoff & von Neumann (1936) and defended by Hilary Putnam (1975) (who later gave it up), that for a simple realist interpretation of quantum mechanics we should reject the distributive laws for conjunction and disjunction. The result is what Birkoff & von Neumann call quantum logic.¹⁴ Putnam argues that the case is just like that of finding that physical space is non-Euclidian, concluding that “[l]ogic is as empirical as geometry” and that “[w]e live in a world with a non-classical logic” (1975, p. 184). We cannot get into the details of Putnam’s proposal here, but an objection often raised against proposals like it is precisely that it is not clear what sense could be made of the analogy between logic and empirical theories.¹⁵ It is one thing to define a system in which sentences of the form ‘ $\phi \wedge (\psi \vee \chi) \leftrightarrow (\phi \wedge \psi) \vee (\phi \wedge \chi)$ ’ are not logical truths, but another to say that we have thereby rejected a distributive law. The problem is illustrated by Quine himself in *Philosophy of Logic* (1986). Considering a case of a deviant logician who proposes that contradictions can be true, he writes:

They think they are talking about negation, ‘ \sim ’, ‘not’; but surely the notation ceased to be recognizable as negation when they took to regarding some conjunctions of the form ‘ $p, \sim p$ ’ as true, and stopped regarding such sentences as implying all others. Here, evidently, is the deviant logician’s predicament: when he tries to deny the doctrine he only changes the subject. (p. 81)

So too with the apparent rejection of the distributive laws, we might say that when someone who maintains that they are not logical truths, their notation of ‘ \vee ’ and ‘ \wedge ’ have ceased to be

¹³ Empirical defence of deviant logic is rare in the literature. Usually, the defence proceeds by what is traditionally seen as a priori argument, as in Dummett 1983 for intuitionistic logic or Priest 2006 for dialethic logic.

¹⁴ As a sidenote, the law of excluded middle is valid in quantum logic. So, the revision suggested by Birkoff & von Neumann, despite also concerning the simplification of quantum mechanics, is not the same revision that Quine alludes to.

¹⁵ For a detailed overview and criticism of Putnam’s proposal, see Berger 2011. The criticism, which also addresses Quine’s idea of the revisability of logic, is pieced together from Kripke’s unpublished lecture notes which have been expanded upon more recently by Romina Padró (2015). From Padró’s work, the criticism has come to be known as the adoption problem. In short, it is that the adoption of a logic presupposes logic and is thus impossible. Although it is, I think, a good criticism of confirmational holism being able to subsume logical beliefs, it is not where the heart of the matter is. The adoption problem shows that we are, as it were, logically bound to our logic but does not explain *why* we are. The final chapter of this thesis hopefully sheds some light on it.

recognizable as expressions of disjunction and conjunction since *those* are characterized by truth-tables which validate the distributive laws.

We can put the worry in the form of an objection for Devitt to respond to. Quine says that to restore an equilibrium between the totality of our beliefs and incoming experience we can reject “statements of logical connections” between our beliefs, that is, we can reject our logical *beliefs*. But what the subject believes about the logical connections between beliefs is one thing, what the logical connections are is another. Whether incoming experience calls for revision in beliefs or whether revision of some beliefs calls for a revision of others does not depend on whether the subject *believes* that a revision is called for. It depends on whether it really is called for. And it is called for when there is an *inconsistency* between what is experienced and what is believed. If a subject believes that $p \rightarrow \neg q$ and that p and their observations suggest that q , there is an inconsistency between experience and the totality of their beliefs. In response, they might give up either of their existing beliefs or adopt an assumption concerning why, despite appearances, $\neg q$. Giving up the *belief* that $\neg(p \wedge \neg p)$ in no way alleviates matters, nor does it help to adopt the belief that $p \wedge \neg p$. Giving up or adopting a logical belief thus seems to do nothing to restore the equilibrium between incoming experience and the totality of the subject’s beliefs. The analogy with the point against deviant logicians is that in the objection we are appealing to inconsistency in just the way we are appealing to a certain construal of conjunction and disjunction when responding to the quantum logician. In a way, we might be said to assume the very thing that the target of our criticism is rejecting. But, if we cannot rely on *logic* to assess the beliefs of others, what *can* we rely on? This is the conundrum that Devitt should have an answer to.

There is what looks to be a response that we can piece together from Devitt’s discussion. First, Devitt maintains a richer notion of what faces the “tribunal of experience”. Rather than taking it to be a totality of beliefs as Quine does, Devitt speaks of “evidential systems” which consist not only of beliefs but also “include dispositions to respond selectively to perceptual experiences and to infer according to certain rules” (2005, p. 108). Among them are dispositions to infer according to deductive inference rules like *modus ponens*, that is, from ‘ $p \rightarrow q$ ’ and ‘ p ’ to ‘ q ’, as well as, perhaps, dispositions to avoid contradictions. Devitt agrees that for a person who has an evidential system like this to be justified in believing what they do, it makes no difference whether they believe the rules to be valid. What matters is that they infer according to the rules and that the rules *are* valid. The role that Quine assigns to logical beliefs is assigned by Devitt to inferential dispositions.

Inferential dispositions are not the sort of things that could be knowledge. They are ways of doing things. But, Devitt adds, a subject can “stand back from her arguments and ask some epistemological questions”, that is, they can assess how good their evidential system is. And the belief that the system is good, Devitt suggests, is plausibly seen as “supported by the empirical success” of the evidential system, whatever the details of the system, just as “someone afloat on a boat may not know the methods by which it was built but, noting its seaworthiness, infers that the methods, whatever they were, are good” (p. 109). Similarly, we can note the seaworthiness of our evidential systems and hence of our inferential dispositions. Our inferential dispositions reflect inference rules which in turn could be formulated as propositions that we may come to believe are true. So, it seems that holism can after all account for the justification of logical beliefs in terms of how they explain incoming experience. Logical beliefs help us account for experiences of our evidential systems being good.

To get a better sense of the response, let us consider an example of what sort of experiences are being explained by logical beliefs. We should note first that the experiences in question cannot be those that the system itself explains. The goodness of the system hangs on those experiences being explained already. The belief that our evidential system is good should explain something else. There are candidates for such experiences. For example, I realized only now—though I could have realized it at the time—that I was able to find my keys when coming home yesterday. Finding the keys must have involved quite a bit of (implicit) reasoning, including reliance on *modus ponens* and the law of non-contradiction. So, my evidential system, along with *modus ponens* and the law of non-contradiction must be pretty good. So, $(p \rightarrow q) \wedge p \rightarrow q$ and $\neg(p \wedge \neg p)$.

We must be careful about what is being explained and what does the explaining. The truth of ‘ $(p \rightarrow q) \wedge p \rightarrow q$ ’ is not explained by my finding the keys. This would put the logical truth in the role of what is given by experience and my belief that I found the keys in the role of what explains it. Rather, my finding the keys must be explained partly by the truth of ‘ $(p \rightarrow q) \wedge p \rightarrow q$ ’. If the proposition wasn’t true, I wouldn’t have found my keys, or at least would have struggled quite a bit more. This, it seems, is Devitt’s proposal concerning how confirmational holism can serve as an account of justification for logical beliefs. It does address the worry we raised previously, when we noted that the subject who rejects the law of non-contradiction is still faced with recalcitrant experience. With Devitt’s introduction of evidential systems, we can also consider cases where the subject simply doesn’t have the requisite inferential disposition. A virtue of Devitt’s proposal is that we can still say that the subject is

faced with recalcitrant experiences. From our own point of view, we can recognize that the inferential dispositions that the subject has are no good. This can still be an objective matter and not simply a dogmatic attitude that we (perhaps inevitably) have. We can, in other words, rely on our logic in assessing others. Moreover, Devitt's proposal counts such a subject as unjustified in holding the beliefs that they do, since the evidential system that the subject has is no good.

But there is a problem at the heart of Devitt's proposal. My finding the keys *is not* explained, not even partly, by the truth of ' $(p \rightarrow q) \wedge p \rightarrow q$ '. Whatever explanation we provide for my finding the keys we can write as many logical truths to it as we like, or we might write none. What explains my finding the keys will concern my habits of taking the keys with me when I leave the building, and my keeping them in the same place when I'm out. And it will plausibly concern my beliefs and my reasoning about where the keys might be. At this point it is tempting to add: and the reasoning, where deductive, should be valid. But *that* adds no more to the explanation than 'either it is raining, or it is not raining'. The explanation will be made no better and no worse by it. We should note that the same problem does not arise for rules that Devitt calls "ampliative" (2005, p. 109). An example of an ampliative rule would be 'when you see smoke, infer that there is a fire'. To explain my success in getting around in the world, such a rule certainly does need to be mentioned. But there is also a clear sense in which the world would be different if the rule was no good. For rules of deductive reasoning, there isn't. The allure of Devitt's proposal is there only when we consider the evidential system as a whole without inquiring into which rules are actually reflected in the dispositions which belong to the system. When we separate out the contribution of each rule to the explanation of incoming experience, we find that the deductive inference rules must be *assumed* to be just like the ampliative ones. But then confirmational holism, as Devitt has proposed it, does not give us the best explanation of our logical knowledge. It gives us no explanation at all.

III

Is A Priori a Way of Knowing?

When we say that we can know logical truths independently of experience, what are we taking the ‘can’ to concern, exactly? On one disambiguation, we are saying that we have a way of knowing things independently of experience that happens to be suitable in logic. On another, we are saying that logical truths are of the sort that knowledge of them does not call for experience. These are different conceptions of what a priori is characteristic of. Let us call the first a *way of knowing* conception and the second a *propositional* conception of the a priori.

Proponents of either conception can help themselves to the terminology of the other. Someone who takes a priori to be characteristic of a way of knowing might define a priori propositions, roughly, as propositions whose truth can be known a priori, whereas someone who construes a priori as characteristic of propositions can define an a priori way of knowing as what is involved in knowing a priori truths. But this does not bring the two conceptions any closer to each other. It only makes it harder to tell them apart. Still, there is a point to it, at least for the proponent of the propositional conception. To explain what the distinction between a priori and a posteriori knowledge consists in, we should explain what the difference between justification in the two kinds of knowledge is. If a proponent of a propositional conception is to provide this explanation, they *need* to have a distinction between kinds of justification. Williamson (2013) introduces some helpful labels for this. In each approach, we can distinguish between the *primary* and the *secondary* classification. We can say that on the way of knowing conception, the a priori-a posteriori distinction is primarily a classification of ways of knowing which is then invoked to define a secondary classification of propositions. On the propositional conception, it is the other way around. But while the secondary classification need not play any role in the way of knowing conception, it is of central importance in the propositional conception if the conception is to give us an epistemological theory at all.

We noted the distinction between the way of knowing and the propositional conception at the start of our investigation, though not quite in these terms. In this chapter, we will explore the distinction in more detail. We will consider how the two differ in approach and what results we can hope for from each. Of the two, the approach driven by the way of knowing conception of the a priori is much more popular in contemporary epistemology. In fact, the other approach is often seen, though mistakenly, as not a proper epistemological account at all. But we will see that, if anything, things are the other way around—that an epistemologically significant

distinction between a priori and a posteriori knowledge can be drawn only with the propositional conception. This will set the stage for the chapters to come where we assume the propositional conception in a defence of the apriority of logical knowledge. Against the way of knowing approach, we will initially consider an argument from Williamson (2013) which aims to show that the a priori-a posteriori distinction does not “carve at the epistemological joints”. But we will find fault in the argument. So, we will develop a different argument for the same conclusion which we will base on some concerns voiced by Albert Casullo (2015).

1. A Priori as a Way of Knowing

The way of knowing conception of a priori, that a priori is a special way of knowing that is independent of experience, was explicitly introduced into contemporary epistemology by Saul Kripke in *Naming and Necessity* (1980). Noting the ‘can’ in Kant’s characterization of something being a priori when it can be known independently of experience, Kripke says: “That means that in some sense it’s *possible* . . . to know this independently of any experience. And possible for whom? For God? For the Martians? Or just for people with minds like ours?” (pp. 34–35).

It might be best therefore, instead of using the phrase ‘*a priori* truth’, to the extent that one uses it at all, to stick to the question of whether a particular person or knower knows something *a priori* or believes it true on the basis of *a priori* evidence. (1980, p. 35)

Since then, the way of knowing conception has become characteristic of a certain methodological approach to developing an account of the a priori-a posteriori distinction. The approach is familiar in contemporary epistemology, at least since Gettier’s (1963) paradigm-shifting paper.

The aim of the approach is to adequately spell out what it is to know something independently of experience. The theorist considers examples of what are intuitively cases of such knowledge. They might consider logical knowledge, mathematical knowledge, conceptual or metaphysical knowledge, moral knowledge, etc., but it is not enough to specify the cases solely by what is known. Since the focus is on knowing, we need to select concrete examples of first-hand knowledge. A case of knowing that Fermat’s Last Theorem is indeed a theorem is no good because (for almost everyone) it is based on testimony. Knowledge that $2 + 2 = 4$ is not ideal, either, because hardly anyone remembers what it is like to have this

knowledge for the first time. But this part of the project is often left implicit, since there is already a shared conception of which cases matter and which do not, at least in rough outline. The appearance that justification in those cases is independent of experience is then, with caution, assumed to be right. Otherwise, there would be little point to try to characterize those cases as cases of a priori knowledge. The theorist then tries to find a general characterization of justification which applies to all the relevant cases of knowing and fits the initial sketch of a priori knowledge as knowledge where justification does not depend on experience. The adequacy of a given characterization is measured, on the one hand, by whether it is in line with the initial sketch of what a priori knowledge is, and on the other, by whether it classifies all and only the apparent cases of a priori knowledge as a priori. Both standards are, of course, open to scrutiny. The first standard decides whether the proposal is a contender for an account of a priori knowledge, and the second decides whether it is a good contender. It is thus characteristic of the way of knowing approach that the account is justified *bottom up* by how well it captures intuitions about specific cases.¹⁶

An account of a priori justification that is developed following the approach is an attempt to characterize justification of a specific kind. It is an attempt to sort cases of justification into those where justification is a priori and those where it is a posteriori. Because of this, the way of knowing conception presupposes that cases of justification can, at least in principle, always be identified independently of whether they are cases of a priori or a posteriori justification. Only if cases of justification are independently given could we make sense of sorting them. Justification, thus understood, is a *genus* in which a priori and a posteriori are *species*. It is possible, given the way of knowing conception, to decide whether a subject is justified in believing what they do and thus whether they have knowledge, while leaving it open whether their knowledge is a priori or a posteriori. Questions like ‘Does the subject know?’ and ‘Why is it that the true belief of the subject is knowledge?’ are genus questions. They do not call for a way of knowing distinction between a priori and a posteriori knowledge. Only an answer to ‘*How* the subject knows?’ concerns the distinction. This is what earns the way of knowing conception its name.

¹⁶ A distinction between a bottom-up and a top-down approach is also introduced by Williamson (2013). But by this, Williamson has in mind only the order of *introduction*, namely whether the characterization is motivated by the examples or whether the characterization is developed independently and then checked against the examples. I use the labels ‘bottom-up’ and (later) ‘top-down’ to characterize the order of *justification*. It does not matter how one come by the characterization. What matters is whether the characterization is right. In the bottom-up approach, this question is settled by considering examples.

The flip side of construing the a priori-a posteriori distinction as a distinction between two ways of knowing is that the distinction will be topic neutral. Even if we built our characterization of the a priori on examples of logical knowledge alone, the characterization itself will still be of a way of knowing and not of what is known. Because of this, a way of knowing account by itself cannot settle which things, if any, can be known in the specified way. An answer to that question needs to be pursued separately. If we consider the account itself as the only constraint, then *any* proposition could be known a priori as well as a posteriori. Whatever limits there are on what is a priori or a posteriori knowable will depend on the nature of truths. It is for this reason that a secondary classification of propositions defined in terms of a primary classification of ways of knowing is unlikely to be worthwhile. When we use our characterization of the distinction between ways of knowing itself as the classifier, then the range of a priori truths will be the same as the range of a posteriori truths. The only way we can get an interesting classification of propositions is if we let the features of propositions determine the classification. But our classification will then be tracking human cognitive abilities regarding those features, not the features themselves. For example, to get a mutually exclusive distinction between a priori and a posteriori truths, we might define a priori truths as truths which can be known a priori and a posteriori truths as truths which can be known a posteriori but not a priori.¹⁷ Truths about current weather could then be said to be a posteriori and not a priori in the sense that humans are able to know such truths by experience only. But there are Martians with different ways of getting at truth for whom truths about current weather are not a posteriori. The Martians may know them a priori or in some way that is neither. Once we have given a positive characterization of what it is to know something a priori there will be space for that too.

There are distinctive ways in which only a way of knowing approach to the a priori-a posteriori distinction can be criticized. One may fail to consider some cases that are relevant or may consider cases which are not relevant. One may also just get the characterization of the cases wrong. For example, in response to George Bealer's (1998) proposal that a priori justification is mediated by a kind of seeming that is intellectual rather than perceptual, Williamson has complained: "I am aware of no intellectual seeming beyond my conscious inclination to believe Naïve Comprehension" (2007, p. 217). There is also the danger of stepping over the line of the initial sketch of what a priori knowledge is. For example, in response to Crispin Wright's (2004) proposal that belief in the truth of basic laws of logic is

¹⁷ A secondary classification like this is proposed by Williamson (2013, p. 293).

default rational, enjoying “a type of rational warrant which one does not have to *do any specific evidential work* to earn” (p. 174), Casullo objects that “[t]he traditional concept of a priori justification . . . is not negative”—that it “requires that a priori justified beliefs *have* a particular type of justification,” and not merely that “they *lack* a particular type of justification” (2015, p. 2720). Debates over all such matters are signals of a way of knowing conception of a priori.

We can also see a way of knowing conception in play in a kind of response to general criticism of the idea of a priori knowledge, like the one we saw stem from confirmational holism. Since a way of knowing account assumes that the cases it is built on are genuine cases of a priori knowledge, it is always undermined by such criticism and cannot be appealed to in response to it. In response to empiricism—that all knowledge depends on experience—Laurence Bonjour (1998) argues that “the rejection of any sort of *a priori* justification leads inexorably to a severe skepticism” (2001, p. 625). What he does not do is infer the existence of a priori knowledge from the existence of rational insight, his own account a priori knowledge. He does not do so because an account based on a way of knowing conception, like his own, implies the existence of a priori knowledge only because it assumes it.

A clear example of an account of a priori knowledge based on a way of knowing conception is Philip Kitcher’s (1980). Kitcher is explicit about his assumption that a priori is a way of knowing, stating that “[a] priori” is an epistemological predicate. What is *primarily* a priori is an item of knowledge” (p. 3). He then adopts from Kant the characterization of a priori knowledge as “knowledge absolutely independent of all experiences,” (p. 4) and, in line with our model, he sets as his goal the spelling out of this sketch.

I shall try to show that Kant’s definition can be clarified, and that the concept of a priori knowledge can be embedded in a naturalistic epistemology. (ibid.)

Kitcher is also explicit in assuming that the a priori-a posteriori distinction is a distinction of species within a genus.

Leaving the task of specifying the conditions on warrants to general epistemology, my aim is to distinguish a priori knowledge from a posteriori knowledge. (Kitcher 1980, p. 8)

Kitcher’s first step is to propose the following positive characterization in place of Kant’s negative one: “X knows a priori that p just in case X has a true belief that p and that belief was produced by a process which is an *a priori warrant* for it” (p. 8). The task is then “to distinguish a priori warrants from other warrants,” that is, from the empirical ones (ibid.). Although Kitcher

does not give us an explicit list of examples, he has examples in mind by which to measure the success of his result.

A clearheaded apriorist should admit that people can have empirical knowledge of propositions which can be known a priori. (p. 6)

Obviously, any theory which implies that I can know a priori that I am seeing red (when, in fact, I am) would be suspect. (p. 15)

Given the account Kitcher arrives at, an a priori warrant is a process of such a type that under any variation of experiences (“given any life,” as he puts it (ibid.)) sufficient to ensure a capacity to have the belief in question 1) some process of that type could produce the belief, 2) and if it did, the belief would be warranted (justified), and 3) also true (p. 10). Given the account, a priori knowledge depends on some process or other that meets the three conditions regardless of experiences. There may be many processes that fit the bill. For example, a process of deriving a conclusion according to deductive inference rules is plausibly of this sort whereas seeing a coffee mug in front of you is not. But nothing in the account itself says that logical or mathematical truths can be known by such processes whereas truths about coffee mugs cannot be known by them.

Unlike Kitcher, BonJour (1998) develops his account by giving us an explicit list of examples (ibid., §4.2). Among them is a case of assessing the validity of inferring that David ate the last piece of cake from two premises: that it was either David or Jennifer who ate the last piece, and that it was not Jennifer. Reflecting on it, BonJour finds that “understanding” the three propositions is enough to “see or grasp or apprehend directly and immediately . . . that there is no way for the premises to be true without the conclusion being true as well” (p. 105). This sort of apprehension, BonJour suggests, “purports to be nothing less than a direct insight into the necessary character of reality” (p. 107).

Such a rational insight, as I have chosen to call it, does not seem in general to depend on any particular sort of criterion or on any further discursive or ratiocinative process, but is instead direct and immediate (pp. 106–107)

In particular, BonJour suggests that rational insight “does not seem to depend on experience . . . , thus allowing it to be the basis for *a priori* justification and *a priori* knowledge” (p. 107). So, like in Kitcher’s account, we are offered a spelling out of the idea of a priori knowledge being independent of experience. BonJour then clarifies the notion of rational insight. Although

the initial examples “[f]rom an intuitive standpoint” suggest that rational intuition is infallible (1998, p. 107), he notes that other examples suggest that it can misfire.

It is as clear as anything philosophical could be that the claim of infallibility . . . is false and completely indefensible. There are simply too many compelling examples of propositions and inferences that were claimed to be objects of rational insight, and hence to be justified *a priori*, but that subsequently turned out to be false or mistaken. (p. 111)

The account we are offered is thus that a priori knowledge is knowledge that depends, not on experience, but on a fallible rational insight into things necessarily being a certain way.

BonJour is less explicit than Kitcher about treating the a priori-a posteriori distinction as one of species within a genus. However, as BonJour says, rational insight not depending on experience *allows* it to be the basis of a priori justification. It does not show it to be such a basis. We must attribute this further role to it. Like Kitcher, BonJour does not say what this role consists in exactly, only that rational insight seems to be playing it—that it “seems to provide an entirely adequate epistemic justification for believing or accepting the proposition in question” (p. 107). But whichever way we spell out the role, it will not be distinctive of a priori justification. What is distinctive about a priori justification is already captured by invoking rational insight.

It is also not immediately obvious that BonJour’s account is topic neutral. On Kitcher’s account, no feature of propositions or truths is mentioned at all. This is deliberate on Kitcher’s part, as he aims to develop his account “without closing the case against the contingent a priori” (1980, p. 9). BonJour, on the other hand, characterizes rational insight partly in terms of necessity which he construes as truth “in any possible world or situation” (1998, p. 106).¹⁸ But it would be a mistake to infer from this that BonJour’s proposal is restricted to propositions of some specific class. Rational insight is meant to be how we come to *learn* about “the necessary character of reality”. We must have the means of learning prior to what we learn. Otherwise, there would be little point to the means. At most, rational insight may involve an *appearance* of necessity. This is clear from BonJour’s emphasis on fallibility. Given BonJour’s proposal,

¹⁸ The construal in terms of truth in all possible worlds is liable to lead to a reading of BonJour’s notion of necessity in terms of the semantics of propositions of the form ‘ $\Box\phi$ ’ in modal logic. For example, Casullo (2003, pp. 15–17) presents a series of objections to BonJour’s account on that basis. And it may suggest that unlike Kitcher’s account, BonJour’s cannot capture cases of contingent a priori. But this is an uncharitable reading of BonJour. As we saw, he also defines validity in terms of his notion of necessity. A proposition like ‘ $p \rightarrow$ actually, p ’, though it is contingent when necessity is understood on the model of ‘ $\Box\phi$ ’, still follows from no premises and is thus still necessary in BonJour’s sense. See Bostock 1988 for a discussion of the sort of necessity that is in play here.

if one “sees clearly and after careful reflection” that the rain is a necessary feature of reality (perhaps on Midsummer Eve in Estonia), then they have rational insight. They would then be a priori justified in believing that it is raining. This falls short of knowledge only because, presumably, it is not raining necessarily.

2. A Priori as a Kind of Truth

When we disambiguate the claim that we can know logical truths independently of experience to mean that logical truths are of the sort that knowledge of them does not *call* for experience, we get a very different approach to a priori knowledge. The a priori-a posteriori distinction is then primarily a distinction between kinds of proposition, not ways of knowing.

This is often misunderstood by those who hold the way of knowing conception. For example, in Casullo’s (2003) comprehensive taxonomy of accounts of a priori knowledge, accounts based on the propositional conception are classified into two kinds: those that appeal to necessity and those that appeal to analyticity (p. 11). Similarly, in his criticism of the a priori-a posteriori distinction, Williamson (2013) considers the possibility that the a priori-a posteriori distinction might be construed in terms of the necessary-contingent distinction, noting that “Kripke taught us not to read the epistemology of a truth off its metaphysical status” (p. 298).¹⁹ But when assuming the propositional conception of the a priori we do not have in mind the necessity or the analyticity of propositions. We have in mind the *apriority* of propositions. We treat it as a feature of a proposition whether knowledge of its truth calls for experience. Like Kant, we construe the a priori-a posteriori distinction as a separate distinction alongside the analytic-synthetic and the necessary-contingent distinction.

The propositional conception can never lead to an account of the a priori-a posteriori distinction that is topic neutral. To even formulate the conception, we need already to specify a class of propositions. We chose to focus on logical truths, but depending on interest, we may have chosen something else, perhaps mathematical truths or moral truths or, indeed, truths about current weather (if we thought that a promising route). But we must specify some propositions to fix the scope of the ensuing investigation. An immediate consequence of it is

¹⁹ There is a difference between Kripke’s and Williamson’s points. Kripke warned against *conflating* the two distinctions whereas Williamson has in mind a construal of one distinction in terms of the other. The justification for the two points is still the same, that the distinctions do not line up.

that the results we arrive at, whatever they are, do not generalize beyond the class of propositions we selected at the start. There will be no guarantee that an account of the a priori status of logical knowledge will be the same as that of mathematical knowledge or moral knowledge. It is of course another matter, if we manage to reduce, let us say, mathematical truths to logical truths. We would then not be expanding the scope of the account but demonstrating that mathematical knowledge was in its scope all along.

The goal of a propositional approach is to explain why knowledge of truths in the selected class of propositions does not call for experience. The approach proceeds primarily by a study of the propositions in question. Examples of what are intuitively cases of a priori knowledge play no role in this, other than in the initial set up where we choose a scope for our investigation. But this role is of a kind with that played by the apple falling on Newton's head in the ensuing theory of gravity. A propositional account does not assume that the apparent cases of a priori knowledge are genuine but is instead meant to demonstrate it in an illuminating way. The approach driven by the propositional conception is thus a *top-down* approach with respect to individual cases. The appearance that knowledge in individual cases is a priori is justified by appeal to the general account. The general account itself is given an independent defence based on an investigation of the nature of propositions.

To explain why knowledge of logical truths does not call for experience, we need to demonstrate two claims about logically true propositions. First, we need to show that because of what logical truths are like, they cannot be known by experience, that any experience involved in knowing them can at most play a merely enabling role. And second, that they can be known. The first claim leaves open the possibility that logical truths simply cannot be known to be true. The second excludes it. Both claims need to be demonstrated by appeal to features of logical truths, that is, the 'can' in neither claim may reflect the cognitive or psychological abilities of subjects. Since our aim is to find an explanation and not merely a demonstration, there are some constraints on what form a defence of the two claims can take. Reductio arguments, for example, though they demonstrate the truth of their conclusions, tend to have little explanatory value.

Experience either is or is not called for in knowledge of a given truth. Thus, given the propositional conception, every truth is either a priori or a posteriori, and no truth is both. As we noted at the start, to provide an account of the a priori-a posteriori distinction based on the propositional conception, we also need to introduce a secondary distinction between a priori

and a posteriori justification. The need for it is captured in an objection by Casullo to accounts based on the propositional conception (it is presumably meant to apply generally, not just to accounts based on necessity or analyticity).

There is a general reason for regarding them with suspicion. The analysandum in question is epistemic. It is a type of justification. An informative analysis, however, should highlight what is distinctive about such justification. An analysis in terms of necessity or analyticity highlights what is distinctive about the propositions so justified rather than the justification itself. Hence, it will fail to be informative. (2003, p. 12)

Casullo's talk of analysis and analysandum is misleading. Whatever "analysis" we arrive at, it will be the result of an attempted explanation. Casullo is still right in saying that the explanandum for an account of a priori knowledge is epistemic. An informative account "should highlight what is distinctive" about justification in a priori and a posteriori knowledge. A propositional account meets this requirement not by directly characterizing justification but by defining kinds of justification in terms of a distinction between propositions. A priori and a posteriori justification get defined, respectively, as justification in knowledge of a priori truths and justification in knowledge of a posteriori truths. What is distinctive about the two kinds of justification will then be highlighted by appeal to why knowledge of a posteriori truths calls for experience but knowledge of a priori truths does not. So, while accounts based on the way of knowing conception have little theoretical use for their secondary classification, in accounts based on the propositional conception, the secondary classification is what eventually gives us an answer to what a priori and a posteriori justification consist in.

By introducing the distinction between kinds of justification in terms of the distinction between propositions, a propositional account treats the justification that a given belief can enjoy as fully determined by the kind of proposition whose truth is believed. Since all propositions are divided without gap or overlap into a priori and a posteriori propositions, there is also no gaps or overlaps in what kind of justification a given belief can enjoy. Beliefs in the truth of the same proposition cannot be justified a priori in one case and a posteriori in another. There is an objection that this is liable to give rise to. As Kripke puts it, "[s]omething may belong in the realm of such statements that *can* be known *a priori* but still may be known by particular people on the basis of experience" (1980, p. 35). For illustration, Kripke introduces a case of a computing machine delivering the result that a given number is prime. We believe the result "on the basis of our knowledge of the laws of physics, the construction of the machine, and so on" (ibid.), even though one could also know this by calculation. Another

example is given more recently by Williamson (2013, p. 292). The proposition ‘ $289 + 365 = 654$ or there are cable cars in Switzerland’, Williamson says, can be known empirically by a person who infers it from the truth of ‘there are cable cars in Switzerland’, and it can be known a priori by a person who infers it from the truth of ‘ $289 + 365 = 654$ ’. The problem with all such examples is that it is only by assuming the way of knowing conception that we could see them as cases of the same truth knowable both a priori and a posteriori. Given the propositional conception, how a subject acquires knowledge makes no difference to whether they are justified a priori or a posteriori and thus to whether they are justified at all. What this shows is that justification itself must be understood in a different way when we assume the propositional conception of the a priori-a posteriori distinction, not as a matter of what the subject does, as Kripke and Williamson understand it, but as a matter solely of the truth in question. For example, some truths may be appropriate to believe only when validated by experience while others may require something else or nothing at all. It might still be possible for one to rely on experience when learning those truths, but this reliance on experience would not make the truths any more appropriate for belief than they already are.

Since a propositional account introduces kinds of justification by defining them in terms of a distinction between propositions, it also does not have room for a neutral notion of justification. The distinction between kinds of justification is not arrived at by dividing an independently given pool of cases of justified belief. So, unlike a way of knowing account, a propositional account does not construe a priori and a posteriori justification as two species within a genus. It construes the distinction instead as one between *two genera*. A neutral notion of justification can still be introduced in a propositional account. Justification in general can be defined as what a priori and a posteriori justification have in common. But it will not be possible to decide in a given case whether a subject is justified in believing what they do and thus whether they have knowledge while leaving it open whether the subject is justified a priori or a posteriori in holding the belief. What justification in a priori and a posteriori knowledge consists in will be construed differently.

A sketch of a propositional account of the a priori-a posteriori distinction is offered by Frege in *The Foundations of Arithmetic* (1980, orig. 1884). Frege characterizes the a priori-a posteriori distinction as follows:

For a truth to be a posteriori, it must be impossible to construct a proof of it without including an appeal to facts, i.e., to truths which cannot be proved and are not general, since they contain assertions about particular objects. But if, on the contrary, its proof can

be derived exclusively from general laws, which themselves neither need nor admit of proof, then the truth is a priori. (*FA*, §3)

Thus construed, it is a classification of truths in terms of the types of proof they afford. A truth is a priori if and only if there is a proof of it from general laws that “neither need nor admit of proof”, and it is a posteriori if and only if it any proof of it involves a particular fact as a premise.

Which proofs a given truth affords is determined by the conditions under which the proposition is true. Given a contemporary conception of (semantic) proof, there is a proof of a conclusion from some assumptions when the assumptions are true, and their truth is demonstrably sufficient for the truth of the conclusion. On this conception, there is a proof, for example, from the assumption that $1 = 1$ and $2 = 2$ to the conclusion that $1 = 1$. Frege speaks of proofs in a more restricted sense. What Frege has in mind are proofs where the truth of the assumptions also *explains* the truth of the conclusion.

The aim of proof is, in fact, not merely to place the truth of a proposition beyond all doubt, but also to afford us insight into the dependence of truths upon one another. (*FA*, §2)

The answer to why $1 = 1$ is not that $1 = 1$ and $2 = 2$. What explains why $1 = 1$ is presumably the reflexivity of identity—the general law that for every x , $x = x$.²⁰ On the other hand, an explanation to why it is raining, though it will involve general laws, will also include particular facts concerning the location in which the rain occurs. And many of the general laws invoked in the explanation, for example, those that capture regularities in weather patterns, are in turn explained partly by further particular facts. So, on Frege’s account, it is an a priori truth that $1 = 1$ but an a posteriori truth that it is raining, and the characteristic difference between them is in what shape an explanation of their truth must take.

Along with the sketch of a distinction between truths, Frege gives us a distinction between kinds of justification that reflects it.

When a proposition is called a posteriori or analytic in my sense, this is not a judgment about the condition, psychological, physiological and physical, which have made it possible to form the content of the proposition in our consciousness; nor is it a judgment about the way in which some other man has come, perhaps erroneously, to believe it true;

²⁰ In *The Basic Laws of Arithmetic*, where Frege sets out to pursue the project outlined in *The Foundations of Arithmetic*, the reflexivity of identity is explained in terms of more primitive laws. See Frege 1964, pp. 112–113 for the proof.

rather, it is a judgment about the ultimate ground upon which rests the justification for holding it to be true. (*FA*, §3)

The explanation of why a belief in a given case is justified and thus why it is knowledge is, on Frege's account, just an explanation of why what is believed is true. And whether knowledge in that case is a priori or a posteriori depends on the shape that this explanation takes. If the truth of what is believed is explained solely by general laws of the requisite sort, then the belief is justified a priori, and if it is explained partly by particular facts, the belief is justified a posteriori.

A consequence of Frege's construal of justification is that justification in general does not concern the subject who holds the belief but only the subject-matter of what they believe. The question whether the subject is justified is "assigned, if the truth concerned is a mathematical one, to the sphere of mathematics" (*FA*, §3).

The problem becomes, in fact, that of finding the proof of the proposition, and following it up right back to the primitive truths. (*ibid.*)

Justification of a mathematical belief, on Frege's account, is taken to be a matter of mathematics, justification of a meteorological belief a matter of meteorology, etc. Given the picture Frege is offering, knowledge is much easier to achieve than contemporary epistemologists would be comfortable with. If the subject's belief is true, then they have knowledge when there *is* a proof of the right sort, regardless of whether they or anyone else has *found* such a proof. If there is an explanation to why it is raining (as surely there is), then anyone who believes that it is raining has knowledge. Of course, this cheap knowledge is of little comfort, since, if the explanation is not found, it will also not be found why the subject knows. What is difficult, given Frege's account, is not the acquisition but the recognition of knowledge.

One feature of Frege's sketch of the a priori-a posteriori distinction that has struck some as odd is that it involves no mention of experience. This has led Tyler Burge (2000) to complain that "[s]ince any modern notion of apriority seems necessarily tied somehow with justificational independence of experience, Frege's omission is, strictly speaking a mischaracterization of the notion of apriority" (p. 22). A problem with this assessment is that Frege explicitly assures us in a footnote that by the terms 'a priori' and 'a posteriori' he aims "only to state accurately what earlier writers, Kant in particular, have meant by them" (*FA*, §3). Frege could not have missed the explicit and repeated appeal to experience in Kant's

characterization of the distinction. In line with our model of the propositional approach, what Frege is giving us is an account of *what it is* for knowledge to depend on experience. He explicitly rejects the idea that empirical justification is to be understood in terms of the occurrence of experience.

If we call a proposition empirical on the ground that we must have made observations in order to have become conscious of its content, then . . . [w]e are making a psychological statement, which concerns solely the content of the proposition; the question of its truth is not touched. In this sense, all Münchhausen's tales are empirical too; for certainly all sorts of observations must have been made before they could be invented. (*FA*, §8)

On Frege's account, for the justification of a belief that it is raining to depend on experience just is for the explanation for why it is raining to concern particular facts. The term 'empirical', on Frege's account, is characteristic of particular facts, of what there is *to* observe or experience. Empirical justification thus understood is still associated with the occurrence of experience in one sense. To see *why* the subject is justified in holding their belief, we would need to look for a suitable proof of what they believe and check whether the assumptions of the proof are true. This will involve observation if the assumptions involve statements of particular facts. So, we may say that, on Frege's view, empirical justification is of the sort that can only be recognized by observation.

Frege's construal of the a priori-a posteriori distinction does not allow for topic neutrality. To develop an account based on Frege's sketch, we would first need to select a class of propositions to check whether they can be proved solely from general laws of the required sort. This is what Frege himself sets out to do later in *The Basic Laws of Arithmetic* (1964, orig. 1893), where he aims to demonstrate that mathematical truths can be proved solely from logical axioms. No intuitive examples of a priori knowledge are being used to legitimate either the sketch or the mature account. Instead, if a class of truths can be shown to follow solely from general laws of the required sort, it will be *explained* why, in believing those truths, we have a priori knowledge. In line with our model, the explanation has two parts. The knowledge is a priori because of the absence of particular facts in the proof of the truths known, and it is still knowledge because the general laws which explain the truths do not themselves need a proof, hence the general laws must "neither need nor admit of proof". The distinction between kinds of justification, as Frege draws it, is not the result of qualifying a neutral notion of justification in two different ways. It is not a distinction between species within a genus, like it is in way of knowing accounts. Given Frege's account, to find whether the subject is justified one needs to

find a suitable proof. There is thus no space for answering the question ‘does the subject know?’ while leaving it open whether the subject’s knowledge is a priori or a posteriori. One cannot find a proof without finding what assumptions the proof has.

There is a problem with Frege’s account. One issue that is sometimes pointed to is the notion of proof that Frege employs. For example, Burge notes that “[a]ny truth can be ‘proved’ within some logical theory, in the usual modern sense of the word ‘prove’” (2000, p. 16). The problem Burge hints at is that proof in Frege’s sense is something quite esoteric that requires a philosophical account of its own. Although it is true that explanation is not officially built into the modern sense of proof and that it calls for philosophical clarification, this does not mean that we could not make perfectly good sense of a distinction between explanatory and non-explanatory proofs. This is attested by the fact that Frege’s notion of proof is still widely assumed by mathematicians today. For example, mathematicians Gower & Neilson (2009) have stressed that “for mathematicians, proofs are more than guarantees of truth: they are valued for their explanatory power, and a new proof of a theorem can provide crucial insights” (p. 879). A genuine problem in Frege’s account concerns the general laws that “neither need nor admit of proof”. It is intuitively clear how a general law might not admit of proof, but what does it mean for a law not to *need* one? Frege does not tell us. The laws that Frege has in mind, as he makes clear in *The Basic Laws of Arithmetic*, are laws of *logic*.²¹ So, Frege’s construal of the a priori-a posteriori distinction, even if it can be applied in explaining the apriority of mathematics, can tell us nothing about the apriority of the laws of logic. Moreover, it is natural to see Frege’s proposal that a priori truths are those whose truth is explained solely by the laws of logic as depending on the implicit assumption that the laws of logic are a priori. So, for a general account of the apriority of logic, we need to turn elsewhere. Still, Frege’s account can serve as our illustration of how the a priori could be understood primarily as a characteristic of truths and how this can lead to an understanding of what justification in cases of a priori knowledge consists in.

²¹ If we replace the problematic Axiom V with Hume’s Principle—that there is a one-one function from *F*s into *G*s iff the number of *F*s is the number of *G*s—which is generally not seen as a law of logic, then, for Frege’s project to work for mathematics, a separate question will need to be addressed, namely whether Hume’s Principle is a priori, and if so why.

3. Williamson's Insignificance Challenge

We have now considered in general terms the two ways of construing the a priori-a posteriori distinction. In this section, we will see why the approach driven by the way of knowing conception is based on a mistake. We will take inspiration for our argument from a different argument given by Timothy Williamson (2013).

Williamson's target is an a priori-a posteriori distinction that is "primarily a classification of specific ways of knowing" (p. 292). Williamson argues that the distinction thus understood, though a real distinction between two kinds of knowledge, is not epistemologically significant, "like a taxonomy of plants and animals based only on colour" (p. 291). Williamson is not clear about what he takes epistemological significance to consist in. Colours of plants and animals, even if they are not a good basis for a taxonomy, are clearly biologically significant. To learn why a tree has red leaves, you ask a biologist. What Williamson plausibly has in mind may be put in this way. Epistemology, though described simply as a philosophical study of knowledge, is really aimed at explaining what it is to have knowledge as opposed to a mere true belief. Although Williamson prefers not to speak of knowledge in terms of justification, he cannot have a problem with our *definition* of justification as the difference of knowledge and mere true belief. To say that a distinction between the a priori and the a posteriori way of knowing is not epistemologically significant is then to say that it is not a distinction between kinds of justification. I will read Williamson as aiming to demonstrate this.

Williamson's strategy is to compare "what would usually be regarded as a clear case of a priori knowledge with what would usually be regarded as a clear case of a posteriori knowledge" and demonstrate that "the epistemological differences between the two cases are more superficial than they first appear" (p. 294). The general conclusion is then that the a priori-a posteriori distinction does not reflect a genuine epistemological difference, or, as Williamson puts it, that it "does not cut at the epistemological joints" (*ibid.*). The cases to be considered must be clear as opposed to borderline since the absence of significant differences between two borderline cases would not undermine the significance of the distinction in general. Viruses are borderline between alive and not alive, but this is hardly comforting for anyone else.

To develop the argument, Williamson introduces what he refers to as "normal cases" of knowing (1) and knowing (2) (p. 295):

- (1) All crimson things are red.

(2) All recent volumes of *Who's Who* are red.

Since both propositions are true, the question whether one knows them to be true concerns justification. Considering the cases, Williamson suggests that ...

normal cases of knowledge of (1) are clearly a priori, because by definition crimson is just a specific type of red, whereas normal cases of knowledge of (2) are clearly a posteriori, because it takes direct or indirect experience of recent volumes of the British work of reference *Who's Who* to determine their colour (ibid.)

Williamson then constructs two hypothetical scenarios to illustrate that it is possible to know (1) and (2) by the same cognitive process. The subject, Norman, learns the relevant concepts not by having them explained to him in terms of other concepts but by ostensive means. Things are pointed out to him while he is told 'this is red', 'this is crimson', 'this is a recent volume of *Who's Who*', etc. As he goes on to apply the concepts in judgment, he gets feedback on his performance from observations. In this way, experience calibrates Norman's skills in judging whether something is red or crimson or a recent volume of *Who's Who*. It becomes more and more likely that when Norman makes a judgment, his judgment is true. Once sufficiently competent, he can apply these skills "offline" in imagination (p. 295). In one of the scenarios, Norman imagines a sample of crimson, judges that it is red, and concludes (1). And in the other, he imagines a random recent volume of *Who's Who*, judges that it is red, and concludes (2).²² Being conceptually competent, Norman is likely to get things right when going through these imaginative exercises. Because of this, Williamson suggests, Norman's two judgments should thus amount to knowledge. And so, Williamson concludes:

The problem is obvious. . . . [T]he cognitive processes underlying Norman's clearly a priori knowledge of (1) and his clearly a posteriori knowledge of (2) are almost exactly similar. If so, how can there be a deep epistemological difference between them? But if there is none, then the a priori-a posteriori distinction is epistemologically shallow. (pp. 296–7)

²² A side remark: When Norman studies a sample of crimson, his direct conclusion can only be that crimson is a *kind* of red. From this it follows that all crimson things are red. But noticing that a recent volume of *Who's Who* is red, no matter how randomly the book is selected and whether it is selected in imagination or off the shelf, will not warrant the generalization that all recent volumes of *Who's Who* are red. If Norman knows the generalization in both cases, he must know them in different ways. Williamson is presumably assuming a parallel with mathematical proofs that begin with a random selection of a number or other mathematical object (as in 'let $n \in \mathbb{N}$ be chosen arbitrarily'). That style of proof requires some generalizations like axioms to be assumed in the background. It does not work for books.

In other words, since Norman comes to know both truths in essentially the same way, despite his knowledge being clearly a priori in one case and clearly a posteriori in the other, the a priori-a posteriori distinction cannot be epistemologically significant.

That something is off with Williamson's argument may be seen from that fact that, in glossing the two cases, he suggests that the cognitive processes that Norman relies on involve experience in a way that is "more than purely enabling but less than strictly evidential" (p. 298) making Norman's knowledge of (1) and (2) neither a priori nor a posteriori. Williamson takes this to be compatible with his previous claim that Norman's knowledge is clearly a priori in one case and clearly a posteriori in the other by supposing that the cases are clear when classified intuitively but are neither a priori nor a posteriori when classified by direct appeal to the criterion of justification by experience. Casullo (forthcoming) has rightly pointed out that the argument cannot be read that way.

As Casullo notes, Williamson "does not specify what constitutes a "normal case" of knowledge of (1) or what constitutes a "normal case" of knowledge of (2)" (forthcoming, p. 7). This matters because Williamson's argument is meant to be an objection to an a priori-a posteriori distinction that is "primarily a classification of specific ways of knowing" (2013, p. 292). The clear cases of a priori and a posteriori knowledge that the argument requires must be individuated not by the truths known but by *how* those truths are known. That is why Williamson needs to talk about "normal cases" of knowing (1) and (2) instead of simply calling knowledge of the two truths, respectively, a priori and a posteriori without more ado. But, as Casullo points out, given the way of knowing approach, "a normal case of knowing p is some particular way of knowing p" (forthcoming, p. 7). So, we need to ask which particular ways of knowing are being classified as clearly a priori and clearly a posteriori in the two Norman scenarios. However, in the two scenarios, Norman knows (1) and (2) in the same way, clearly. Whatever the normal ways of knowing the two truths are, the imaginative way that Williamson introduces for the scenarios *replaces* them. And we can no longer say that Norman has clearly a priori knowledge in one scenario and clearly a posteriori knowledge in the other.²³

²³ Despite the failure of the argument, Williamson's Norman cases still challenge an assumption often implicitly held by proponents of the way of knowing approach, that the a priori-a posteriori distinction is jointly exhaustive of cases of knowledge. But it is hardly a serious criticism. Knowledge by testimony and knowledge by inference are ways of knowing, as is knowledge by osmosis (which is how we know, for example, that there are no orchids on the Moon). Williamson's knowledge by imagination seems to be just one more item on the list. Given that Williamson's goal, as it is clear in *Philosophy of Philosophy* (2007), is to explain the acquisition of modal knowledge in terms of imagination, this may well be enough for his purposes.

The obvious explanation of what is going on in Williamson's construal of the argument is that the classification of knowledge of (1) as clearly a priori and knowledge of (2) as clearly a posteriori is tracking a distinction between propositions. The phrase 'normal cases' means nothing if we have already specified a way of knowing for both truths. One might diagnose the confusion by reading Williamson as getting side-tracked by the secondary distinction of propositions that the way of knowing conception affords. But a closer reading suggests a different confusion. As we have seen, Williamson does say something about the normal ways of knowing that he has in mind. He identifies knowledge of (1) as a clear case of a priori knowledge "because by definition crimson is just a specific type of red" (2013, p. 295). This is a characterization of a proposition not a way of knowing. A proponent of a way of knowing account *couldn't* see this as an explanation to why knowledge of (1) is a priori. Moreover, Williamson consider knowledge of (2) a clear case of a posteriori knowledge "because it *takes* direct or indirect experience . . . to determine" the colour of a book (ibid., *emphasis added*). This too is a characterization of a proposition. It is to construe the proposition itself as demanding experience for its verification. Williamson thus assumes the propositional conception when classifying the cases as clearly a priori and clearly a posteriori, and then switches to the way of knowing conception to note a lack of difference in ways of knowing in the two cases. The result is an objection to neither.

I suspect that Williamson is not alone in having this confusion. It is common in contemporary epistemology to agree with two ideas about apriority. On the one hand, we tend to agree with Kripke that apriority is a matter of intuitive details of individual cases of knowing. On the other hand, we also wish to agree, with some reservations, with the traditional framework in which mathematical and logical truths among others are said to be a priori. We think that the two ideas can be made compatible by relying on the Kripke-inspired way of knowing conception to define a secondary classification of propositions. But as we have seen, the secondary classification that the way of knowing conception affords can do no more than provide a traditional appearance. That is, it allows for us too to talk about a priori truth. But what we then mean by the term is very different from what people used to mean by it before Kripke. If we then still go on to theorize about the apriority of mathematics, logic, conceptual truths, etc. we are bound to keep running into confusions of just the sort that we found in Williamson's argument.

4. The Insignificance of Ways of Knowing

Williamson's argument against the way of knowing approach fails, but what he aimed to demonstrate is still true. At least so I will argue in this section. But first, we should clarify the intended conclusion. Williamson aimed to show that the a priori-a posteriori distinction, understood as a distinction between ways of knowing, is "epistemologically shallow" (2013, p. 297), meaning that "epistemologists would do better to avoid the distinction between the a priori and the a posteriori in their theorizing, because it distracts them from deeper similarities and differences" (p. 291). But finding that differences between two kinds of knowledge are not relevant can itself be an epistemological discovery. And it can take considerable effort to explain why exactly they are not relevant, even if they appear to be. There is thus no guarantee that epistemologists can or should avoid talking about a shallow distinction. The distinction between testimonial and first-hand knowledge is an example. Beliefs that we form about matters we have observed for ourselves seem to be in much better epistemic standing than beliefs we form about matters we have heard about only from others. In fact, it is initially hard to see how the latter could have any epistemic standing at all. When we then realize that almost all our beliefs are formed in the second way, it seems to severely undermine much of what we took ourselves to know. It is a task of epistemologists to explain why this is not so, why the differences between first-hand and testimonial knowledge are not so significant as they appear. We may say that the distinction, though epistemologically significant, is hopefully not *epistemically* significant—that it does not make a difference to whether the subject has knowledge. It is in this sense that I will argue that the a priori-a posteriori distinction, understood as a distinction between ways of knowing, is not significant.

The way of knowing approach proceeds by spelling out the initial characterization of the a priori-a posteriori distinction as a distinction between knowledge that depends on experience and knowledge that does not. The result will be an epistemically significant distinction just in case it is motivated by considerations that pertain to the difference between knowledge and mere true belief. In spelling out the distinction, there are two questions that we need to answer before we even get to the mystery of a priori justification, namely 'what is experience?' and 'what is it for knowledge to depend on it?'. We have a rough answer to the second question. For knowledge to depend on experience is for experience to play a role in the explanation of why the subject has knowledge rather than a mere true belief. So far, we have said nothing about experience itself. But it plays just as big of a role in classifying cases of knowledge into a priori and a posteriori.

If introspective awareness of one's own mental states was not an experience, then knowledge that depends on this awareness would be a priori and not a posteriori, given the way of knowing conception. Similarly, if rational insight was simply the product of a further sense, no different from the more familiar five senses, then knowledge gained by it would be a posteriori and not a priori. So, how we answer the question 'what is experience?' makes a difference to how cases of knowledge get classified. A need to address the question has been recognized by proponents of the way of knowing conception. BonJour (1998) notes that "[w]hile the general intent of this appeal [to experience] seems clear enough at first glance, it turns out to be surprisingly difficult to delineate precisely" (p. 7). Casullo (2003) provides a detailed analysis of the difficulty. The goal is to specify all and only those cognitive processes or states which can play a justifying role in a posteriori knowledge. There are broadly speaking two ways to go about this. One is to offer an exhaustive list, and another is to provide a general characterization of what a cognitive process or state must be like to count as experience.

Among the problems that Casullo notes in the first approach is that an explicit list "has little explanatory value" (2003, p. 149). Cases of knowledge would count as either a priori or a posteriori, depending on what we have included on the list. But this classification would be as arbitrary as the list itself. There would be "no rationale for dividing sources of justification into two fundamental categories: the experiential and the nonexperiential" (ibid.). In other words, the a priori-a posteriori distinction would be pointless. To have an a priori-a posteriori distinction that is motivated, we need to explain why a given cognitive process or state is experience. This requires a general characterization of experience.

To proceed in a search of a general characterization, Casullo suggests fixing a paradigm case of experience which is "sense experience in its various forms" (p. 150), that is, visual experience, auditory and olfactory experience, etc. The task, he says, is then to "identify some feature of sense experience that is possessed by all, and only, the other relevant types of experience" (ibid.). Casullo considers a list of initially plausible candidates, finding fault in each one.²⁴ Among the candidates is the proposal that "[t]he cognitive states associated with each of the five senses have a characteristic phenomenology by virtue of which they are readily

²⁴ For Casullo's critique of characterizations of experience, see Casullo 2003, §6.2 and Casullo 2015. Among his targets is the account that BonJour (1998, p. 8) adopts after noting the difficulty of saying what experience consists in. The proposal is a causal theory of experience; roughly, that what is characteristic of experience is that it is caused by its object. Casullo's (2015, p. 2711) objection is that the distinction between a posteriori and a priori knowledge will then effectively be drawn by what can take part in causal relations and what cannot, that is, between concrete and abstract objects, which, given the way of knowing conception, is not an option.

identifiable,” (2015, p. 2709) and thus that we might take this characteristic phenomenology to be the mark of all experience, not just of the familiar sense experience. An idea like this seems to be echoed, for example, in Bealer’s (2002) distinction between intellectual and experiential seemings. This proposal is of interest first because Casullo’s response to it fails to bring out its real flaw, and second because this flaw generalizes to other accounts of experience.

Casullo rightly notes that for a phenomenological account of experience to succeed “there must be some phenomenological feature common to all and only those cognitive states that are experiential sources of justification” (2015, p. 2709). Casullo objects that not even all cases of standard sense experience, like auditory and visual experience, seem to have phenomenological features in common, and that it is thus unlikely that a phenomenological account is right. But the main problem with the account is not that it is probably wrong. The problem is that a shared phenomenology, even if there was one, would not be the reason why experience is playing a justifying role in our a posteriori knowledge. All experiences would have this phenomenology regardless of what is experienced. And if we used the resulting phenomenological notion of experience to spell out the a priori-a posteriori distinction, we would get a significant phenomenological distinction but not an epistemic one.

In response to failures to explain what experience consists in, Casullo (2003) suggests taking ‘experience’ to be a natural kind term, which would imply that the nature of experience is to be “uncovered by empirical investigation rather than a priori reflection” (p. 159). The promise of turning to cognitive psychology for answers is that, in the best-case scenario, “[w]e will have a characterization of nonexperiential processes that tells us not only *which* processes are nonexperiential but also *why* they are nonexperiential” (ibid.). Casullo is right about this. But what he fails to consider is just what the imagined proponent of the phenomenological account fails to consider. We do not just need a criterion by which to classify ways of knowing. We need an *epistemic* criterion by which to do so. If cognitive processes are classified as experiences or as not experiences by standards of cognitive psychology, then cases of knowledge are classified as a priori or as a posteriori by those *same* standards of cognitive psychology. That is, an answer to why knowledge in a given case is a priori or a posteriori will come from cognitive psychology, not from epistemology. And the a priori-a posteriori distinction would thus be a psychological distinction. By this, I do not just mean that there would be some psychological element involved in it. I mean that the distinction would be *purely* psychological. The only respect in which the distinction concerned epistemology would be the respect in which a priori and a posteriori knowledge are the *same*. The task of

epistemology would be to explain why the subject is justified in holding the belief in both cases *despite* the differences.

A proposal similar to Casullo's is made by Williamson (2013), that cognitive psychology "will have much to offer the epistemologist's attempt to overcome philosophical prejudices and classify according to deeper and less obvious similarities and differences" (p. 309). What Williamson has in mind is the way in which Norman gains knowledge in the hypothetical scenarios by using his cognitive abilities offline in imagination. But just like Casullo's a priori-a posteriori distinction, Williamson's distinction between online and offline knowledge is a psychological distinction only. We can see this from how Williamson himself describes the cases. Having given a detailed characterization of how Norman forms the belief by imagination, Williamson wraps it up with: "Since his performance was sufficiently skilful, background conditions were normal, and so on, he thereby comes to know" (p. 295). What we have been shown, in other words, is that *despite* Norman's peculiar use of imagination in arriving at a belief, we can still count this belief as knowledge by the usual epistemological standards. So, the significance of the Norman cases is this. Previously, epistemologists may have thought that it matters epistemically whether the person observes directly or simply imagines—that the person can have knowledge in the first case but not in the second. What Williamson shows is that it does not matter. The distinction, though psychological, "does not cut at the epistemological joints" any more than the corresponding a priori-a posteriori distinction does.

The problem that we have identified is not specific to the phenomenological account of experience that Casullo objects to or the cognitive psychology account he goes on to endorse. It derives from the way of knowing conception itself. As we saw earlier, an account based on the way of knowing conception is an attempt to sort cases of justification into cases of a priori justification and cases of a posteriori justification, or more generally, into cases with different species of justification. Because of this, it assumes that justification as such—the genus—can be characterized independently of the classification. Only if we have some grasp of cases of justification in general can we go on to sort them. The problem we have identified is a direct consequence of it, namely that questions concerning justification, 'Why does the subject have knowledge rather than a mere true belief?', can always be answered without invoking the distinction. We can always decide whether the subject is justified in believing what they do while leaving it open whether experience is playing a role in this. The role of experience matters

only in an answer to a different question, namely to ‘How does the subject know?’. And that question, as Popper emphasized, is not a question of justification.

In the way of knowing accounts, what motivates the classifications are the examples of what appear to be cases where knowledge does not depend on experience. And the question that a proponent of a way of knowing account aims to answer is how the belief could be knowledge despite this appearance. They may try to answer it by trying to resist the appearance, or they might accept the appearance and spell out the details of the case to show that experience not playing a justifying role in the case does not really undermine the knowledge. The a priori-a posteriori distinction thus construed never does epistemological work. Rather, epistemological work is always done on it.

Williamson’s analogy between epistemology and biological taxonomies is thus an apt one. The taxonomies of plants and animals are taxonomies of *data* that call for explanation. The aim of the biologist is to find a theory in which the distinctions in the taxonomy no longer play a role. Occasionally, however, a life form is found that cannot be accounted for in the existing theory. Archaea were thought to be bacteria of a sort, until it became clear that a separate branch must be introduced to the evolutionary tree to accommodate them. In this way, our background assumptions about life itself changed, and a distinction was introduced into the theory. This is the sort of distinction that a proponent of the propositional conception of the a priori has in mind.

In response to the objection to the epistemic insignificance of ways of knowing, it may be tempting to turn to specific theories of justification, like reliabilism or evidentialism. Reliabilism, as it is taken in a nutshell, is the view that a subject is justified in believing what they do just in case their belief is obtained by a reliable process, a process that is more likely to produce true beliefs than false ones. A proponent of the way of knowing conception might add that acquiring beliefs by experience is a reliable process and that the involvement of experience is thus still relevant to justification. But this is to take advantage of an ambiguity. As noted already by Ramsey, the first to explore reliabilism, ...

the word ‘process’ is very unsatisfactory; we can call inference a process, but even then unreliable seems to refer only to a fallacious method not to a false premiss as it is supposed to do. . . . Perhaps we should say not . . . obtained by a reliable process, but . . . formed in a reliable way. (Ramsey 1929, p. 258)

Even a valid inference is unreliable when it proceeds from false premises. The way the process itself unfolds is not what its justifying power consists in. What Ramsey emphasizes is that it is not any process that accounts for justification but only the fact that a belief is produced in a way that tends to lead to true belief. The specific mechanisms of belief production are irrelevant. It is rational to give up a belief after finding that it is the result of taking hearsay at face value, not because it was hearsay that led us to hold the belief, but because hearsay is not a reliable way of forming beliefs. As Ramsey puts it, “we should remain certain if, and only if, we thought our way reliable” (ibid.), and not, we may add, if we thought our way experiential. When Goldman’s (1979) work brought reliabilism to the mainstream, he had this point clearly in view, describing justification as “a function of the reliability of the process,” not as a function of a process that is reliable (p. 10). This is because what matters is reliability. The process itself, whether it be experience, a priori insight, or an imaginative exercise, always factors out.

The situation is analogous with evidentialism, the view that a subject is justified in believing something just in case they have evidence for its truth. Just as in reliabilism it is not the process that justifies but its reliability, so in evidentialism, it is not the *having* of evidence that justifies but the *evidence*. Given evidentialism, justification “is determined by the quality of the believer’s evidence for the belief” (Conee & Feldman 1985, p. 15). Experience, understood as a way of acquiring or having evidence, factors out as epistemologically irrelevant. What matters is the evidence or the quality of the evidence.

There is a potential response to this. Although ways of knowing do not play a role in deciding whether the subject is justified in holding the belief, we do not merely want a yes or no answer to the question. We want an explanation. That is, we want to know *why* the subject is justified. Different things will play a role in the explanation. We would need to investigate what justification the subject has; for example, how good is their evidence or how reliably their beliefs are formed. Investigation into the nature of justification in general will also be relevant at this point. But in addition to those questions, we may also ask where the justification is coming from, what is the source of the justification. Ways of knowing will be relevant to our understanding of the subject’s justification at this point. For example, to explain why my belief that the Universe is expanding is knowledge rather than a mere true belief, one will appeal to the observations of the scientists which confirm this. Here, it is not the observing but what is observed that is said to confirm the truth of the belief, but learning about the observation still seems to give some insight about why I am justified in believing that the Universe is expanding.

I suspect that this response is confused but will not try to argue the point here. The immediate problem with it is that even if some insight is gained in individual cases from learning how the subject knows it does not mean that any *classifications* of ways of knowing will have epistemic significance. There will still be no explanatory value in dividing justification into species. In particular, there would be no explanatory value in a distinction between a priori and a posteriori ways of knowing.

IV

Alice on Logical Knowledge

Alice laughed. “There’s no use trying,” she said: “one *ca’n’t* believe impossible things.”

“I dare say you haven’t had much practice,” said the Queen.

(Carroll, *Through the Looking-Glass*, p. 174)

The a priori-a posteriori distinction is primarily a distinction between kinds of proposition, or so we will assume from now on. As we have seen, an inevitable limitation of this approach to the a priori-a posteriori distinction is that we cannot assume from the outset that all propositions traditionally taken to be a priori are sufficiently alike to warrant a unified account of them. The aim of the present work is to demonstrate the apriority of logical knowledge. So, from this chapter onwards, our focus will be on logic only.

In this chapter, we will build on the view held by Alice in her exchange with the White Queen to give a sketch of how logical knowledge could be distinct from non-logical knowledge by virtue of the nature of propositions of logic, of logical truths and falsities. Our aim is to clarify what it would mean for logical falsities to be strictly unbelievable and show how, once properly understood, it implies the apriority of logical truths. A defence of the idea will be postponed until the last chapter.

We will first introduce Alice’s thesis, restricting it to logical belief, and will spell it out by taking further cues from Berkeley and Kant who both agree with Alice. We will see that the thesis is not a statement of any psychological limitation on what we can believe, as the Queen takes it, but a statement of a limit to what there is *to* believe. This we can clarify with the help of a contemporary notion of an epistemic possibility. We will see how the thesis, thus construed, sets logical knowledge apart from all other knowledge and implies that logical knowledge, *if* we are to have it, must be a priori. Finally, we will consider an attempt to demonstrate that we *have* logical knowledge from Alice’s thesis alone. The failure of the attempt is instructive in showing us what more is needed and thus how to proceed in the coming chapters.

1. Alice's Thesis

On the surface, it seems that what is at issue in the disagreement with which we began between Alice and the White Queen is belief in something you *know* to be impossible. How else to make sense of the idea of practicing belief in impossible things? However, as we will see below, it is plausible that the view held by Alice and denied by the White Queen is that one cannot believe things which are *in fact* impossible whether one knows them to be impossible or not. Given this reading, the exchange might seem to a contemporary reader to be a rare occasion where a character of the Looking-Glass world is the one to express common sense, and where Alice is the odd one. For surely impossibilities *can* be believed. As Mark Jago in his defence of the White Queen's view puts it:

We think about impossible things all the time. We can think about alchemists trying to turn base metal to gold and about unfortunate mathematicians trying to square the circle. We can think about what those people tried to do and say, truly, that what they tried to do is impossible. (Jago 2014, p. 1)²⁵

Our focus however is on logic. So, our immediate concern is not with all impossibilities, like 'base metal could be turned to gold' or 'there are round squares', but only with *logical* impossibilities, that is, with logical falsities. The official formulation of Alice's thesis that we will be working with may thus be given as follows:²⁶

Alice's Thesis

It is impossible to believe of a logical falsity that it is true and of a logical truth that it is false.

But even if we restrict the thesis to logic, common sense seems to be against it. To take an example closer to our topic, in his magnum opus *The Basic Laws of Arithmetic*, or *Grundgesetze* for short, Frege aimed to show that logic can serve as a foundation for arithmetic. In June 1902, after Frege had already sent the second volume to print, he received a letter from Bertrand Russell, where Russell shows that the basic principles of Frege's *Grundgesetze* jointly imply a contradiction. The culprit was Frege's Basic Law V, that any concept defines a class. After noting this historical event, Mark Jago asks:

²⁵ The wording of this passage is somewhat unfortunate. It is one thing to think *about* impossible things and another to *think* them. Jago's focus is really on the latter, as is ours.

²⁶ The first to attribute the thesis to Alice was Ruth Barcan Marcus (1983), calling it Alice's view. Though Marcus discusses the more general thesis that *no* impossibility can be believed.

Why would Frege have written the combined 565 pages of the *Grundgesetze*, if he knew at the outset that its core laws were contradictory? (2014, p. 43)

Surely, Frege knew no such thing. He believed that the principles were consistent. And so, Frege believed a logical falsity. And if it could happen to someone like Frege, it can certainly happen to any one of us. Or so the objection goes.

Despite the apparently obvious counterexamples, Alice's thesis has cropped up on numerous occasions in the history of philosophy. (It is for this reason that Carroll likely had the stronger view in mind when writing the exchange between Alice and the White Queen.) For example, we find Alice's thesis endorsed by Berkeley. Having argued that the idea of matter is incoherent, Berkeley goes on to say:

Strictly speaking, to believe that which involves a contradiction, or has no meaning in it, is impossible In one sense indeed, men may be said to believe that matter exists, that is, they act as if the immediate cause of their sensations . . . were some senseless unthinking being. But that they should clearly apprehend any meaning marked by those words, and form thereof a settled speculative opinion, is what I am not able to conceive. This is not the only instance wherein men impose upon themselves, by imagining they believe those propositions they have often heard, though at bottom they have no meaning in them. (*Treatise*, §54)

We may read Berkeley's term 'contradiction' as covering all logical falsities, not just those of the form ' $p \wedge \neg p$ '.²⁷ It is noteworthy that Berkeley says he is "not able to conceive" of someone believing a contradiction and that contradictions "have no meaning in" in a person. We will come back to what to make of this in the next section.

We find the same idea in Kant's *Critique of Pure Reason*. When addressing a topic concerning analytic judgments but before reaching the specific case of analyticity, Kant writes:

Whatever the content of our knowledge may be . . . , all our judgments in general must submit to that general, though only negative, condition that they do not contradict themselves; otherwise these judgments in themselves . . . are nothing. (*CPR*, A150)

The principle of contradiction, Kant adds, "is a criterion of all truth" and thus "applies to knowledge merely as knowledge in general, regardless of its content, and simply declares that the contradiction entirely destroys and annihilates it" (A151). That is, the constraint that one

²⁷ 'Contradiction,' as Berkeley uses the term, covers not only all logical falsities but also what we today would classify as metaphysical impossibilities.

cannot believe contradictory things to be true does not, according to Kant, have to do with anything specific to the mind of a given person (or of the minds of all people), say, a psychological limitation of some sort. It has to do with what it is to think. There is, according to Kant, no such thing as judging a contradiction to be true. That is the sense in which, according to Kant, contradiction “annihilates” judgment. Echoing a view similar to Kant’s, Frege says of arithmetical truths that

we have only to try denying any one of them, and complete confusion ensues. Even to think at all seems no longer possible. (*FA*, §14)

This Frege takes to be a mark indicating that arithmetical truths are logical truths: “Should not the laws of number, then, be connected very intimately with the laws of thought?” Thus, Kant and Frege both seem to agree with Berkeley that they are “not able to conceive” of someone believing or judging a logical falsity to be true.

Entering the 20th century, we find Alice’s thesis endorsed also by Wittgenstein who maintains that

What is thinkable is possible too. (*TLP*, 3.02)

The flipside of Wittgenstein’s claim is that what is not possible is also not thinkable, and thus we again have Alice’s thesis. Through Wittgenstein’s writing, the thesis also survives in the works of logical positivists, as we will see later. And closer to the present day, we find a version of the thesis endorsed by Robert Stalnaker (1984) and Agustín Rayo (2013). This is not to say that all these philosophers held the exact same thesis and for the same reason. Some, like Stalnaker, seem to see this as an unhappy consequence of what they consider to be an otherwise well-motivated theoretical framework. Having argued for identity conditions of the semantic values of propositions (of propositional contents) which imply that necessities are necessarily believed, and impossibilities are impossible to believe, Stalnaker writes:

[S]ince we have an argument to show that the identity conditions are right, as well as examples that seem to show that they are wrong, the proper response is not so clear. (1984, p. 24)

He then proceeds to develop an account logical and mathematical knowledge that would handle the counterexamples without requiring us to give up the identity conditions of propositional contents.

On the other hand, Wittgenstein presents a defence of Alice’s thesis as part of the end-goal of the *Tractatus*, stating that “the aim of the book is to draw a limit to thought” (*TLP*, Introduction). Similarly, Kant makes heavy use of Alice’s thesis, deeming it “the supreme principle of all analytic judgments” (*CPR*, A150). Like Wittgenstein, Kant sees the thesis as reflecting a limit to “judgments in general”. As we saw above, Kant’s focus is on the mind “destroying itself” when trying judge logical falsities—when trying to oppose the principle of contradiction (*CPR*, A151). The significance that both Kant and Wittgenstein see in this limit is epistemic. More specifically, they both see it as a key to the apriority of logic. But to recognize how Alice’s thesis might yield this result, we first need to understand the thesis properly. This is the goal of the rest of this chapter.

2. A Limit to Epistemic Possibilities

To properly address the idea that Alice’s thesis might imply the apriority of logic, we need to be clear about what the thesis says. As the quotes from Berkeley, Kant and others have already indicated, this is not a straightforward matter.

The White Queen’s response to Alice was that whatever limitations there are concerning belief in logical falsities they are psychological limitations, limitations that one could perhaps overcome with sufficient practice. And even if Alice with her human psychology cannot do so, the Queen herself, an anthropomorphic chess piece with sufficiently different cognitive abilities, faces no such problems. But, as was clear from Berkeley’s and Kant’s elaboration, the thesis is not meant to describe any psychological limitations of subjects. The impossibility of believing a logical falsity must be understood to concern logical falsities themselves, not believers or their psychological states.

We can drop the apparent appeal to psychological limitations if we read the thesis as concerning *epistemic possibility*—what could be the case for all we know. What warrants the label ‘epistemic’ for epistemic possibility is not just the appeal to knowledge in this rough characterization. Epistemic possibilities are matters which are left open as opposed to those which are already known. And ‘left open’ here means left open for us to *believe or disbelieve*. In other words, one reason why epistemic possibilities deserve to be called epistemic is that the range of epistemic possibilities is the range from which all knowledge is, in a certain sense, drawn.

Because of the link to belief, we can use the notion of epistemic possibility to define a non-psychological sense of ‘can be believed’. Let us say that

it can be believed that p if and only if it is epistemically possible that p .

Epistemic possibility, like all possibility in contemporary literature, is understood in terms of truth. The general schema for this is:

it is possible that p if and only if, in some specified range of worlds, there is a world in which it is true that p

If we can understand epistemic possibility with the help of the same schema without building any psychological limitations into the resulting biconditional (for example, into the specification of the range of worlds), then we have specified a sense of ‘can be believed’ that is understood solely in terms of truth in a world. Alice’s thesis can then be read as the claim that it is not epistemically possible for a logical falsity to be true or for a logical truth to be false.

Epistemic possibility is commonly assumed to come in two variants. Given one variant, something is epistemically possible just in case it is logically consistent with what is known. (Though, since knowledge implies belief, what is known would first need to be specified to avoid implicit appeal to belief in the definition.) This variant of epistemic possibility is an instance of what, since Smiley (1963), is known as *relative* possibility. A definition of any type of relative possibility employs some set of propositions and defines possibility in terms of logical consistency with the truth of those propositions. For different sets of propositions, we get different types of relative possibility. In the epistemic case, the set of propositions is the set of those propositions that are known to be true. For physical possibility, it is the set of all laws of physics. Because all types of relative possibility appeal to logical consistency, logical possibility is at least as weak as any relative possibility. That is, if something is not logically possible, then it is also not relatively possible since what is logically impossible is also not logically consistent with anything else. So, epistemic possibility understood as a type of relative possibility trivially adheres to Alice’s thesis.

But there is also another variant of epistemic possibility, which, in contrast to the relative variant, we may call *absolute* epistemic possibility. Given some logical falsity ‘ p ’ whose negation is not yet proved, it is not hard to imagine someone who is in the process of deciding whether to try a proof or a disproof utter: ‘it could be that p ’. What the speaker seems to be

conveying is an epistemic possibility—that the truth of ‘ p ’ is left open by what they know, and that the ensuing logical investigation will hopefully settle the matter, resulting in knowledge that p or knowledge that $\neg p$. If we give the possibility claim a standard interpretation in the possible worlds framework, then, supposing that the claim is true, we must conclude that there is a world—an epistemically possible world—in which it is true that p . Since there are no logically possible worlds in which it is true that p , this second variant of epistemic possibility would be weaker than logical possibility. It is not logically possible for ‘ p ’ to be true but it is still epistemically possible.

Given epistemic possibility in the absolute sense, Alice’s thesis is a highly non-trivial claim. In effect, it says that the range of what is epistemically possible in the absolute sense is the same as the range of what is epistemically possible in the relative sense. In other words, the thesis states that *logical* possibility is a limit to what is absolutely epistemically possible. If the thesis was true, we would need to make sense of apparent cases of logical falsities being epistemic possibilities in a very different way, if we could make sense of them at all. It is this non-trivial reading that we want to give to Alice’s thesis. Since it is only the absolute sense of epistemic possibility that we are concerned with, we may reserve the label ‘epistemic possibility’ for it alone and ignore relative epistemic possibility from now on.

Now that we understand the thesis as stating a limit to what is epistemically possible, we need to clarify how there could be a limit like this. Supposing, for the sake of elaboration, that Alice’s thesis is true, what we need to explain is *why* the truth of a logical falsity and the falsity of a logical truth are not epistemic possibilities. Here, we are not yet looking for a defence of the thesis. We need only to elaborate it enough to know what to defend in the coming chapters.

3. Stalnaker’s Limit by Intension?

A natural place to turn to in search for an explanation of why logical falsities are excluded from the range of epistemic possibilities is Stalnaker’s (1984) work on intentionality. It is where we find the most recent and well-known attempt to defend something like Alice’s thesis.

Stalnaker develops his account of intentionality by first sketching an outline of it which he calls “the pragmatic picture” (1984, p. 4). The sketch is all we need for our purposes. “Rational creatures,” Stalnaker says, “are essentially agents” (*ibid.*). His proposal is then to

construe beliefs and other propositional attitudes “in terms of the role that they play in the characterization and explanation of action” (ibid.). Stalnaker gives a rough sketch of what this role consists in.

What is essential to rational action is that the agent be confronted, or conceive of himself as confronted, with a range of alternative possible outcomes of some alternative possible actions. The agent has attitudes, pro and con, toward the different possible outcomes, and beliefs about the contribution which the alternative actions would make to determining the outcome. One explains why an agent tends to act in the way that he does in terms of such beliefs, and attitudes. (ibid.)

The key point for us to take from this is that, given the pragmatic picture,

the primary objects of attitudes are . . . the alternative possible outcomes of agents’ actions, or more generally, alternative possible states of the world. (ibid.)

Belief, in other words, is to be understood in terms of the agent’s relationship to possible worlds. Roughly, to believe that p , given the pragmatic picture, is for ‘ p ’ to be true in every world that the agent is related to. However, since logical impossibilities are *not* possible outcomes of action, no range of possible worlds that might capture a belief of an agent will include logically impossible worlds. This has led Stalnaker to conclude that no agent will ever come out as believing a logical falsity. And so, we appear to have Alice’s thesis. But this appearance is mistaken. The pragmatic picture, *pace* Stalnaker, does not imply Alice’s thesis.

In Stalnaker’s account, we can think of the belief of an agent as a range of possible worlds that the agent is confronted with. But the agent cannot be confronted with the worlds in the way I am confronted with, say, the range of coffee-mugs and glasses in front of me. The range of things in front of me is given to me by the elements of that range. I can count them one-by-one. This is not how one *can* be confronted with a range of possible worlds. To be confronted with a range of possible worlds, one needs to delineate that range in general terms by something that all the worlds in the range have in common. That common feature is that in each the same proposition is true—that is, the same truth-condition obtains in each. And we are given no reason in the pragmatic picture for thinking that logical falsities are not acceptable for this task. The resulting range of worlds will be *empty*, but that just ensures that the logical falsity *is* true in every world in the range. The subject will thus count as believing the logical falsity, given the pragmatic picture. So, we will not really have Alice’s thesis on Stalnaker’s account.

Other related implications that Stalnaker's account appears to have are dispelled in the same way. The problem that Stalnaker is primarily concerned with—what he calls the problem of equivalence—is that whoever believes that p will, on the face of it, also count as believing everything that is logically equivalent to ' p ' since, given their equivalence, they are true in exactly the same possible worlds. But as far as the pragmatic picture is concerned, we can distinguish between the truth-conditions of ' p ' and ' q ', between what each *requires* for its truth in a given world. Each of the two propositions will then be construed as determining a range of worlds independently of the other. Since we need to accept truth-conditions in the account to explain how the subject relates to a range of worlds, we may as well allow truth-conditions to play a role in the individuation of things believed. A belief that p and a belief that q will then not be the same, even if ' p ' and ' q ' are true in all worlds of the same range, since the agent will count as relating to that range in different ways.

One reason why Stalnaker does not explore this route seems to be a confusion about semantics. Stalnaker is working within a contemporary framework where propositional attitudes like belief are construed in terms of the subject's relationship to propositional *contents* (which, in contemporary terminology, are simply called propositions). Propositional contents are commonly construed as characterizing what the subject understands in understanding the proposition. But they are, at the same time, construed as something different from just the condition under which the proposition is true. Propositional contents, in other words, serve as a middle-layer between the syntax of the proposition and what is the case if the proposition is true, that is, the truth-condition. Stalnaker proposes to spell out the middle layer itself in terms of truth, stating that the definition of propositional contents that he defends—that a propositional content is “a function from possible worlds into truth-values” (1984, p. 2)—“has its origin in formal semantics” (p. 1). Such functions are sometimes called *intensions*. To get Alice's thesis out of this notion of propositional content, we would need to identify truth-conditions with intensions. That is, we would need to assume that a proposition is true in some worlds and false in others by virtue of the intension that it has. Only then would there not be a *further* notion of a truth-condition to block Alice's thesis in the above way. However, an intension is not assigned to a proposition arbitrarily. Things must be a certain way in a world for a proposition to be true in that world. And which way things must be is determined compositionally by the expressions that make up the proposition. So, intensions cannot be identified with truth-conditions because they *presuppose* truth-conditions. Truth-conditions are what provide the explanation to why a given proposition is true in a given world.

A general lesson we can draw from Stalnaker's pragmatic picture, given that we eventually want to defend Alice's thesis, is that the thesis must be read as concerning truth-conditions in a compositional sense. The idea of an epistemic space *is* tied to something like intensions. For each proposition, there is a function from epistemically possible worlds into truth-values that the proposition has in the worlds. What the problem with Stalnaker's pragmatic picture shows is that, to get Alice's thesis, a mere appeal to truth in epistemically possible worlds is not enough. The thesis must concern *why* propositions have the truth-values that they have in epistemically possible worlds.

4. 'Logically False Belief' as Nonsense?

Another way that the limit to epistemic possibilities might get drawn is suggested already in the quotes from Berkeley and Kant. Berkeley claimed he is "not able to conceive" of someone having a logically false belief and maintained that logical falsities "have no meaning in" those who are under the illusion of believing them. Kant's point that logically false judgments "are nothing" might be understood in the same way, not merely as a claim that there are no logically false beliefs but that the very idea of one is incoherent. In short, the reason for the absence of logical falsities in the space of epistemic possibilities may be that the very idea of a logical falsity being an epistemic possibility—the idea of a logical falsity being believed—is strictly nonsensical. Or more precisely, what we should deem as nonsensical are sentences of the form 'S believes that *p*', where '*p*' is a logical falsity.

We cannot, however, count sentences as nonsensical at will. For a sentence to be nonsensical, it must not receive a truth-condition in an adequate semantic theory for the language. For example, it is plausible that an adequate semantic theory will be many-sorted, where terms are classified by domains that they take values from.²⁸ In a semantic theory like this, the name 'Caesar' will likely receive a referent from the domain of persons and the expression '... is a prime number' will likely combine with terms that take referents from the

²⁸ One motivation for construing semantics in this way is suggested by Dummett. In short, Dummett (*Pol*, pp. 73–76) notes that we need to distinguish between terms according to the criteria of identity associated with them. The criteria are those by which a speaker can tell whether reference in two different cases is made to the same object—by which they can recognize an object "as the same again" (p. 73). Different names will receive different criteria of identity and will get classified by virtue of that. Personal names and numerical expressions, for example, will receive different criteria of identity. The way one recognizes a number as the same again is surely different from how one recognizes a person as the same again.

domain of numbers. Consequently, ‘Caesar is a prime number’ will not receive a truth-condition, meaning that it will be nonsensical in light of the theory. Similarly, to count sentences which result from an attempt to combine a belief operator with a logical falsity as nonsensical, we must find something that could be responsible for the nonsensicality of the combination.

A logical falsity, Kant maintains, “is always rightly denied” whereas its negation “must necessarily be affirmed” (A151). Thus, Kant maintains that there is only a problem with affirming a logical falsity, not with denying it, or with affirming a logical truth. Assuming that the point about affirmation carries over to belief, this might suggest that the problem concerns only the specific combinations of belief with a logical falsity. We can put the idea in terms of expressions. Let ‘ p ’ be true just in case it is raining. We might read Kant as maintaining that ‘ $p \wedge \neg p$ ’ is meaningful, and ‘S affirms that ...’ is an expression which normally can be completed with a meaningful sentence to yield a new meaningful sentence, but when we combine the two to form ‘S affirms that $p \wedge \neg p$ ’ we get something nonsensical, even though ‘S denies that $p \wedge \neg p$ ’ and ‘S affirms that $\neg(p \wedge \neg p)$ ’ are both meaningful. But, given this reading of Alice’s thesis, we need to explain why the belief operator ‘S believes that ...’ or its analogues like ‘S affirms that ...’ discriminate in this way between logical falsities and other propositions, even though the belief operator and any logical falsity are individually meaningful and syntactically compatible.

One potential reason for the discrimination is suggested by Frege. Frege famously distinguishes between “the apprehension of a thought—thinking” and “the recognition of the truth of a thought—judgment” (*T*, p. 294). We will consider Frege’s notion of a thought in more detail in the next chapter. For now, it is enough to take it as what one understands in understanding a proposition. The value of Frege’s distinction is commonly seen in the insight that there must be room in our cognitive lives for the mere apprehension of a thought, something that does not involve commitment to the truth of the thought. A standard illustration is belief in something of a conditional form. In believing that, when it rains, it is cloudy, I needn’t believe either the antecedent or the consequent. Yet, I must apprehend the thoughts expressed by both. Otherwise, the conditional belief together with a further belief that it is raining could not rationally lead me to believe that it is cloudy. This is what Peter Geach (1965) has dubbed *the Frege point*. But the flipside of the same point is that judgment *does* involve commitment to the truth or falsity of a thought. In judgment, we do not simply operate with the thought but recognize the thought *as true*. The same holds of any propositional attitude where

the subject is committed to something being the case, including belief and affirmation. Given this point about judgment, we can recognize a feature of belief that is not apparent on the surface. To believe or judge or affirm something is to be committed to its truth. We might thus maintain that it is not nonsensical to speak of commitment to the truth of ' $p \wedge \neg q$ ' whereas it *is* nonsensical to speak of commitment to the truth of ' $p \wedge \neg p$ '. This in turn might explain why an attempt to construct a belief-report with a logical falsity as its subclause amounts to nonsense, even though the subclause itself may function in a way that is essentially no different from a proposition that expresses a contingent thought. And so, we would have an explanation to what determines the limit to the space of epistemic possibilities that yields Alice's thesis.

Locating the source of nonsensicality in the implicit involvement of truth in belief reports is not enough, though. We need to explain why the involvement of truth should have this effect. There are two ways we might try to explain it. One is to suppose that 'it is true that ...' combined with a logical falsity simply does not result in a truth-condition in just the way that '... is a prime number' combined with 'Caesar' presumably does not. If a belief report 'S believes that p ' always unpacks as 'S believes that it is true that p ', as Frege's point about judgment might be taken to suggest, we would consequently deem all belief reports with logically false subclauses as nonsensical. The problem with this idea, however, is that it requires us to construe ' p ' and 'it is true that p ' as expressing different thoughts, that our understanding of one is distinct from our understanding of the other. Only if 'it is true that ...' made a contribution to the truth-condition of 'it is true that p ' could it place a semantic constraint on which propositions it combines with to discriminate between logical falsities and other propositions. But 'it is true that ...' *cannot* make a contribution like this.

If an understanding of 'it is true that p ' involves more than an understanding of ' p ', it would be possible to learn that p while remaining clueless as to whether it is true that p . And an investigation into whether it is true that p would be distinct from an investigation into whether p . It would plausibly be something one should do *in addition* to finding out whether p . It would be one thing, for example, to learn about whales and another to learn the truth about whales. Truth would then not be the goal of all investigation but only of the more specific investigation of what is true. And it would be hard to see why it ought to be the goal of *any* investigation. For it would remain inexplicable why we should ever believe the truth of the matter if the matter all by itself was enough. This is a confusion about what truth is. So, an understanding of ' p ' must be the same as an understanding of 'it is true that p ', and thus 'it is true that ...' can make no contribution to the truth-condition of the proposition in which it

occurs. But then, as we noted, it also cannot discriminate between logical falsities and other propositions.

Although the source of nonsense cannot be the combination of truth with a logical falsity, we can also now see why the appeal to truth can in any case not be what the difference between judgment and apprehension consists in. We can switch between ‘ p ’ and ‘it is true that p ’ *wherever* we have an expression of a thought, including in our characterization of apprehension. To apprehend the thought that p is to apprehend the thought that it is true that p . Judgment, on Frege’s account, should be something else. We do not judge the thought that p , or equivalently, the thought that it is true that p . Rather, what we judge is *that* the thought is true. Here, Charles Travis (2018) notes a different distinction that might account for the nonsensicality of the idea of logically false belief.

Representing things *as* being thus and so is to be distinguished from representing them *so to be*. Representing something *to be* so is beyond a thought’s power. In representing things *to be* thus and so, a representer assumes responsibility. A thought cannot do that. (Travis 2018, pp. 52–53)

Travis distinguishes between two senses of ‘represent’. In one sense, a *thought* represents things as being in some specific way and is thus either true or false, depending on whether things are that way or not. But this, Travis notes, is not enough to account for the possibility of a subject being right or wrong. For that, we need to recognize another sense of ‘represent’ (though we need not use the same word for it, surely) in which there is room for a representer—a *subject* who represents things. Representation in this sense is what is characteristic of propositional attitudes like belief where, as Travis puts it, the subject “assumes responsibility” for the truth of the thought. The distinction between the two senses of ‘represent’ is thus a natural way of taking Frege’s distinction between apprehension and judgment. Apprehension involves representation only in the first sense. Judgment involves representation also in the second sense where the possibility of the subject being right or wrong arises. The two senses of representation are closely linked. The subject represents things to be a certain way by invoking a thought which represents them as being that way. So, in representation in the second sense—in judgment—a thought is employed into a specific role. The proposal for making sense of Alice’s thesis could then be that a logical falsity cannot play this role, that there is no such thing as taking responsibility for its truth. There would thus be no judging that that it is true that $p \wedge \neg p$ is true or believing that $p \wedge \neg p$.

There is a difficulty with this proposal too. How is the special role of a thought reflected in the semantics? It needs to be if an attempt to put a logical falsity into this role is to give rise to *nonsense*—the absence of truth-conditions for sentences like ‘S believes that $p \wedge \neg p$ ’. What makes this a difficulty is that the special role of the thought should not, on the face of it, bear on truth or falsity in any way. All issues concerning truth-values are settled by thoughts themselves representing things *as* being one way or another. But then it is not clear how the special role that thoughts can play—their role in explaining how a subject could be right or wrong—could be something for truth-conditional semantics to capture.

The difference between a thought representing and a subject using the thought to represent is, on the face of it, a psychological difference. What the subject assumes responsibility for when representing things to be a certain way is the truth of a thought—things being as they are represented by the thought. On a natural way of understanding this, the special role that the thought plays in judgment is fully explained by its place in the subject’s psychology. The only difference between a thought apprehended and what is judged would then be a difference in how the thought is *treated* by the subject. Once the assumed responsibility is factored out, all we are left is the thought, or equivalently, the truth of the thought. Given this understanding of the proposal, Alice’s thesis would amount to a psychological thesis after all. We would have to locate the reason why logical falsities cannot be believed in the practice of assuming responsibility for the truth of a thought. The nature of logical falsities may play a role in this but only a contingent role. In principle, we could have different practices of assuming responsibility which would allow for belief in logical falsities. Alternatively, if we suppose that there *is* a non-psychological difference between the thought and what we assume responsibility for, then we end up inflating the notion of truth. There would then have to be a difference between the thought that p and the truth of the thought that p , since the latter just is what we assume responsibility for. This would give us absurdities that we noted previously, namely that, in inquiry, one needs to choose between an investigation into whether p and an investigation into whether it is true that p .

We have now covered the attempt to exclude logical falsities from the space of epistemic possibilities by finding something in the combination of ‘S believes that ...’ with a logical falsity that might make the combination illegitimate. We have seen that the result is either an inflated notion of truth or a psychological reading of Alice’s thesis. Neither is desirable. The reason for the dilemma is simple. Judgment can be divided into a psychological component—the act of judging—and a non-psychological component—what is judged. If the reason why

judgment discriminates between logical falsities and other propositions resides in the psychological component, we have a psychological reading of Alice's thesis. So, the reason must reside in the non-psychological component—in what is judged. But if the non-psychological component always involves a thought, the discrimination can result only from something that is added to the thought. The only candidate is truth. And so, we end up with an inflated notion of truth. In the next section, we will consider a way around this dilemma.

5. Truth-Values without Truth-Conditions

One way in which it might be epistemically impossible for a logical falsity to be true is if nothing was required for a logical falsity to be false, that is, if logical falsities were false without anything to *make* them false, so to speak. Since the negation of a logical falsity is a logical truth, the same would hold of the truth of logical truths. In general, propositions of logic, given this proposal, would have truth-values without truth-conditions.

Like with all possibility, we understand epistemic possibility in terms of possible worlds. Something is epistemically possible just in case it is true in an epistemically possible world. But a proposition is true in a world just in case things are a certain way in that world. Which way they need to be is specified compositionally by the proposition. Even if epistemically possible worlds are worlds in a different sense than possible worlds normally understood, there must still be some features also of epistemically possible worlds that determine which proposition is true in which world. Given the proposal we are now considering, there are no such features in the case of propositions of logic. Consequently, there could be no epistemically possible worlds in which a logical falsity is true or a logical truth false. And thus, there would be no believing of a logical falsity that it is true or of a logical truth that it is false. This would be due to the semantics of propositions of logic themselves. Alice's thesis would thus be a consequence solely of the nature of propositions of logic.

There is a problem, though. We saw that the idea of a logical falsity being false regardless of how the world is extends also to logical truths. If logical falsities do not require anything for their falsity, then their negations do not require anything for their truth. If a proposition is indeed true in a world if and only if things are a certain way in that world, we should be forced to conclude not only that logical falsities are not true in any epistemically possible worlds but also that *logical truths* are not. But if it is not epistemically possible for a logical truth to be

true, then one cannot believe and thus *know* of a logical truth that it is true. Logical knowledge, in other words, would be impossible. The assumption that leads to the problem is that the truth of a proposition in a world *requires* that the world have suitable features. But this is just what we rejected when supposing that logical truths might be true without having truth-conditions. When we give up the assumption, we instead get the result that logical truths are *vacuously* true in *every* epistemically possible world. If logical truths are true regardless of which truth-conditions obtain, it does not matter which world we consider. A logical truth will always come out true there. However, the initial difficulty with accounting for the truth of a logical truth in a world is still significant. We have found that there are *two ways* in which a given proposition may be true in a world—either vacuously or because of what the world is like. This is where Alice’s thesis can be seen as having implications concerning logical knowledge.

6. Two Kinds of Knowledge

Our goal is to show that we have a priori knowledge of logical truths. But, as we will see in the next section, Alice’s thesis by itself does not give us this result. To get the result we are after, we also need to see why Alice’s thesis is true. But even without a defence of the thesis, we can see how it sets logical knowledge apart from all other knowledge. For purposes of exposition, let us assume Alice’s thesis under the reading we have developed for it.

A logical truth, as we now assume, is vacuously true in all epistemically possible worlds whereas a non-logical proposition is true in a given epistemically possible world by virtue of some features that the world has, features that may be present in one world but absent in another. In this sense, the truth-value of a non-logical proposition is *answerable* to how things are, whereas the truth-value of a proposition of logic is not. A proposition of logic, we might say, has a truth-value on its own accord. To this distinction between two kinds of propositions corresponds a distinction between two kinds of *belief*. There is belief that is answerable to how things are and belief that is not. This distinction carries over to knowledge. Non-logical knowledge is knowledge of matters of fact, that is, of how things are. Logical knowledge is not. Logical knowledge thus differs from other knowledge not only in subject matter but also in what *kind* of propositional attitude it is. Knowledge that it is either raining or not is of a different kind than knowledge that it is raining. It does not imply a fact known. This is our first

step towards an epistemologically deep distinction between logical and non-logical knowledge.²⁹

A further epistemic implication of Alice's thesis concerns the acquisition of knowledge. If the truth of a logical truth is not answerable to how things are, then no amount of investigation (empirical or otherwise) of how things are will be relevant to learning of a logical truth that it is true. Logical truths would stand in contrast, as Leibniz puts it, to "truths of fact, which are drawn from sense-experience" (*New Essays*, 75). A priori truths, as we characterized them in the previous chapter, are truths the knowledge of which does not call for experience. Alice's thesis thus implies the apriority of logic. If we have logical knowledge, given Alice's thesis, this knowledge must be independent of experience. This, however, is of little comfort since Alice's thesis, as we have characterized it so far, even if it was shown to be true, leaves some significant question unanswered. We have seen that logical knowledge cannot be acquired by investigating the world, but we have not seen how it *can* be acquired. And most importantly, to show that our logical beliefs are a priori knowledge, we need to explain why they are *knowledge* in the first place. So far, we have only a negative characterization of logical knowledge.

7. Knowledge by Alice's Thesis?

We have now seen how Alice's thesis sets logical knowledge apart from non-logical knowledge. But might the thesis already imply that our logical beliefs *are* knowledge? In this section, we will look at one attempt to get this consequence from the thesis. Its failure will help us see what else is needed if we are to get the conclusion we are after.

We find an allusion to the possibility of Alice's thesis implying that we have a priori knowledge of logic in Kant's *Critique*. After noting that the principle of contradiction sets a boundary for cognition—that contradiction "annihilates" cognition—Kant continues:

Nevertheless, a positive use may also be made of this principle [the principle of contradiction], that is, it can be used not merely to banish falsehood and error (insofar as they arise from contradiction), but also to know truth. . . . For the opposite of that which

²⁹ We are assuming Alice's thesis for logic only. Of course, lumping all non-logical knowledge together is an oversimplification. If we extended the thesis, for example, to mathematics or morality, then we would maintain that knowledge in those areas is also special in the same way as logical knowledge is. Of course, we would then have to defend the thesis in its extended sense. And this is not what we have set out to do.

already lies, and is thought as a concept, in our knowledge of an object is always rightly denied (*CPR*, A151)

The flipside of the impossibility of believing logical falsities is that, in logic, one is bound to get things right. Furthermore, we can agree with Kant that we have this infallibility simply by virtue of having the capacity for knowledge—that is, by virtue of an access to a range of epistemic possibilities from which to draw knowledge. But getting something right is not sufficient for having knowledge. There must therefore be more to Kant’s proposal if it is to be a candidate for an account of logical knowledge. The idea seems to be that not only are we bound to get things right in logic but that we are bound to *recognize* that we get them right simply because any logical truth is an epistemic necessity. The proposal, in other words, seems to be that we have access to the limits of the range of epistemic possibilities, so to speak, from the inside.

The idea that there could be this sort inside access to the limits of the epistemic space is natural, once Alice’s thesis has been adopted. Given that there are some things one is bound to get right, it is natural to suppose that there may be a way to detect them—a mark that sets them apart from other propositions. This is perhaps what Descartes had in mind by his proposal that clear and distinct ideas are true. Similarly, by relying on the specific reading of Alice’s thesis we are assuming, one might think that the fact that logical truths are epistemically necessary may be a detectable mark like this. The idea is most clearly invoked by Berkeley in his attempt to prove that there are no mind-independent qualities of matter.

[W]hat purpose is it to dilate on that which may be demonstrated with the utmost evidence in a line or two, to any one that is capable of the least reflexion? It is but looking into your own thoughts, and so trying whether you can conceive it possible for a sound, or figure, or motion, or colour, to exist without the mind, or unperceived. This easy trial may make you see, that what you contend for, is a downright contradiction. (*Treatise*, §22)

Berkeley invites us to investigate, by focusing only on the contents of our minds, whether mind-independent qualities could exist. Our inability to conceive such qualities is then taken as proof that there cannot be any. The way the argument is supposed to run, it seems, is this. To conceive something as being the case is to suppose the world to be a certain way. Since in this hypothetical scenario, you consider the world from your perspective, you thereby try to put yourself in the position of someone who believes the truth of a given proposition. So, we can take the inability to conceive to be a case of being unable to believe. But supposing that

Berkeley is right about our inability to conceive of mind-independent qualities, why should it imply that mind-independent qualities are impossible?

Berkeley's argument seems to appeal to Alice's thesis (as we saw previously, he does maintain the thesis). And the idea seems to be that, *since* we cannot conceive impossibilities, our inability to conceive *something* shows it to be "a downright contradiction." An obvious problem here is that Alice's thesis, even on Berkeley's own construal that we considered above, is a conditional, not a biconditional. It says that, if '*p*' is a logical falsity, then it is not possible to conceive that *p*. It does not also say that, if it is not possible to conceive that *p*, then '*p*' is a logical falsity. Given this, for an argument like Berkeley's to work, we would need to make one further assumption, namely that we have no psychological limitations. If that was so, then the only explanation to our failure to conceive something would be the nature of what we are trying to conceive. The fault could then lie nowhere else. But the assumption that we have no psychological limitations is obviously false.

In addition to our personal biases, both theoretical and otherwise, there are also plenty of propositions whose falsity is clearly inconceivable, even after careful consideration, and yet whose falsity is known to be possible. Euclid's Parallel Postulate is an example. Closer to our topic, Frege's Basic Law V may be another. And there are, of course, also propositions that strike us as neither obviously true nor obviously false. A case in point is the Continuum Hypothesis—that there is no set whose cardinality is strictly between that of the rational numbers and that of the real numbers. These psychological limitations are not just a problem for Berkeley's attempt to detect the limits of the epistemic space from the inside. Any attempt will run into the same problem. A *subjective impression* of a limit to epistemic possibility can only reach as far as our psychological limitations permit.

Despite this problem, we can find the same idea of an internal access to epistemic necessity also in more recent epistemology. We find it, for example, in BonJour's (1998) explanation of why beliefs acquired by rational insight—"a direct insight into the necessary character of reality" (p. 107)—should amount to knowledge.

[S]uch a rational or *a priori* insight seems to provide an entirely adequate epistemic justification for believing or accepting the proposition in question. What, after all, could be a better reason for thinking that a particular proposition is true than that one sees clearly and after careful reflection that it reflects a necessary feature that reality could not fail to possess? (ibid.)

To account for the cases that illustrate our psychological limitations, Bonjour does go on to emphasize that such insight into the necessity of a proposition is fallible. The view he opts for instead is that “it is *apparent* rational insight . . . that provides the basis for *a priori* epistemic justification” (ibid., p. 113). So, on Bonjour’s proposal, one is justified in believing that p by virtue of it appearing to them that it is necessary that p . Since ‘ p ’ may still in fact be false, one might fail to have a priori knowledge that p , but, given the proposal, one is still a priori justified in believing that p . But the problem with Bonjour’s fallibilist version of the idea is the same as the problem with the original infallibilist version. The issue was never the assumption of infallibility. Rather, it was that no amount of conviction *could* be a sign of truth. One cannot read Alice’s thesis the other way around.

The failure of the attempt to establish the epistemic status of our logical beliefs directly by Alice’s thesis shows what we are still missing. To recap, we have seen that the impossibility of belief in logical falsities, when understood not as a psychological limitation but as a reflection on the nature of propositions of logic, implies that the truth of logical beliefs cannot be checked against how things are and thus that logical knowledge cannot depend on experience. In other words, given Alice’s thesis, we cannot have a posteriori knowledge of logical truths since logical knowledge could then not be understood as concerning matters of fact. But this *calls* for a new understanding of logical knowledge. It does not provide it. Although Alice’s thesis implies a non-psychological distinction between kinds of knowledge, it does not yet give us a non-psychological distinction between kinds of justification. There is then a temptation to relax the demands of a propositional conception of a priori knowledge and look for a *way of knowing* that is suitable for logical truths understood as epistemic necessities. But the epistemic necessity of logical truths posited by Alice’s thesis is a consequence of propositions of logic having truth-values *vacuously* in every epistemically possible world. And Alice’s thesis, as stated, does not say what it is to know a vacuous truth. What it says is only that an account of such knowledge is needed to explain what logical knowledge is.

In the following chapters, we will clarify the nature of logical truths which will eventually give us the means not only to demonstrate Alice’s thesis but to explain what gives rise to it. We will then retrace the steps we took in this chapter in spelling out the implications of the thesis but then already with an understanding of why the thesis holds. In short, we will see that the vacuous truth of logical truths is a consequence of the nature of truth. Alice’s thesis, as we have outlined it in this chapter, will then help to spell out the epistemic implications of this result. We will find that our logical beliefs are justified as vacuously as they are true.

V

The Laws of Truth

To know is to know what is true. To attribute knowledge, one must state what is known to be true. Any understanding of the nature of knowledge thus presupposes an understanding of what truth is. Despite this, a closer examination of truth itself tends not to be part of contemporary epistemological investigations. An intuitive grasp of the notion of truth that we evidently all must have is generally assumed to be sufficient. But there is no guarantee that it is sufficient. An investigation into what truth is can reveal oversights in our study of knowledge.

Logic, as Frege describes it, “is the science of the most general laws of truth” (*PW*, p. 128). That is, the laws of logic are the laws of truth. We uncover such laws by a study of what it is to be true, that is, by a study of what truth is. As Frege puts it, “[t]he meaning of the word “true” is explained by the laws of truth” (*T*, p. 290). If an investigation into what truth is—into “the meaning of the word “true””—reveals how the truth of some things is related to the truth of others, then the explanation of why these relations hold is found to reside in the nature of truth itself and not in the features of any specific thing that is true. Logical truths are propositions that are true solely by virtue of the laws of truth. Logical truths are thus peculiar in that their truth is determined by what truth is.

Since knowledge is of what is true, knowledge of logical truths is knowledge of something that pertains to everything there is to know. This makes logical knowledge stand out from all other knowledge. To understand the respect in which this makes logical knowledge special, we must begin with an investigation of the laws of truth, the subject-matter of logical knowledge.

To seek an account of “the meaning of the word “true”,” we need a criterion by which to decide whether any given account of it is adequate. A suitable criterion has been proposed by Alfred Tarski who used it for developing a definition of truth. Tarski’s definition has since then become rooted in the mainstream, at least officially. But we will find some fault in it. The source of the problem, as we will see, is Tarski’s failure to fully understand Frege’s theory of generality that his definition of truth aims to capture. To avoid the same problem, we will spend some time on clarifying Frege’s theory and the notion of a thought that is operating in the background of it. The insights gained from Frege’s work will help us produce a definition of

truth free of the problem. It will have the added benefit of being much easier to work with than Tarski's definition. To have a precise way of talking about logical truths, we will finally define what a logical truth is in light of our definition of truth.

1. Convention T

The mainstream definition of truth, the one found in most logic textbooks today, is developed on the basis of the definition given by Tarski (1944, 1956). In developing the definition, Tarski realizes that the first step of any such project must be to set a standard by which to assess the adequacy of the definition. He gives us two such standards: a material and a formal criterion (1944, henceforth *SCT*). Given the formal criterion, the definition should be consistent. We will take that for granted and will keep to Tarski's suggested restriction for ensuring that the formal criterion is met, namely that the definition must not be allowed to apply to itself. Our focus will be on the material criterion. Following Tarski (1956), we will refer to it as *Convention T*.

Noting that truth may be attributed to things of different kinds, Tarski restricts his definition to the truth of (declarative) *sentences*—grammatical arrangements of elements of a basic vocabulary of words or other signs. Convention T may now be stated as follows: a definition of truth must imply for every sentence '*p*' in the object-language a biconditional of the form

(T) '*p*' is true if and only if *q*

in the meta-language, the language in which we state the definition. We will call (T) the general T-schema. And we will call biconditionals of this form T-sentences. On the left-hand side of a T-sentence, a sentence of the object-language is mentioned. On the right-hand side, '*q*' is either the object-language sentence '*p*' itself or a translation of it. The whole T-sentence states the condition under which the sentence is true, its truth-condition. In the special case where the object-language sentences are included in the meta-language, the general T-schema takes the form

p' is true if and only if *p*

We will call this the homophonic T-schema, and the T-sentences that are its instances we will call homophonic T-sentences. A notable feature of homophonic T-sentences is that they strike

us as *trivial* truths. To use Tarski's (*SCT*, p. 343) famous illustration, given Convention T, an adequate theory of truth in English for English sentences implies the homophonic T-sentence

'Snow is white' is true if and only if snow is white

We will see why the triviality matters.

When the object-language has only finitely many sentences, Convention T can be met easily. We could construct a T-sentence for each sentence in the language by going through its sentences one-by-one. But any interesting object-language will have infinitely many sentences. For example, in English, we can write 'it is not the case that' to the left of any sentence to get a new sentence to which the expression can be attached again to get a further sentence, and so on. But any language that can be specified at all, even if it has infinitely many sentences, will have only a finite basic vocabulary of signs that all its sentences are composed of. To give a definition of truth that adheres to Convention T, we may thus try to specify some features of the signs that make up the basic vocabulary and do so in a way that would imply a T-sentence for any sentence in which they occur. As a result, each basic sign is treated as making its distinctive contribution to the truth-condition of any sentence in which the sign occurs. We will call these contributions the *semantic values* of the basic signs. In this, we follow Tarski who himself calls his conception of truth "the semantic conception" (*SCT*, p. 345).

As Tarski notes, to specify truth-conditions for all sentences by a treatment of the basic signs, we need to define truth *recursively*, by first characterizing the semantic values of the basic signs and then introducing rules that characterize the semantic values of complex expressions in terms of the semantic values of other expressions. Although our immediate reason for recognizing the basic vocabulary and the need for a recursive definition is a technical one—we cannot give the definition of truth otherwise—it also contains a general lesson about the nature of truth. Given Convention T, *any* understanding of truth must cover all the infinitely many sentences of the language. Thus, any understanding of truth presupposes seeing sentences as having compositional structure. This applies as much to the theorist who considers the object-language from the outside as it does to whoever uses the object-language. For illustration, consider a point made by Anscombe.

It is obvious enough that a proposition divides up into parts. It is also obvious that the division is not arbitrary. You cannot divide 'The cat is often drunk' into 'The cat is of' and 'ten drunk', although each part could be significant (*IWT*, p. 75)

Although it certainly *is* obvious that ‘the cat is often drunk’ cannot be divided into ‘the cat is of’ and ‘ten drunk’, despite both being meaningful, we also want an explanation to *why* the sentence cannot be divided that way. With respect to the shape of the sentence as it is written down, we certainly *can* divide it that way. We could, for example, cut the paper in half in just the right way. The sense in which we cannot make the division is that an understanding of the truth-condition of the sentence is part of a general understanding of truth, and *that* requires seeing the language as a whole as having a specific compositional structure. In this structure, the truth-condition of ‘the cat is often drunk’ is not determined by the values of ‘the cat is of’ and ‘ten drunk’.³⁰

Semantics as a discipline where Convention T is assumed as a criterion of adequacy is now commonly known as formal semantics. Any definition of truth which adheres to Convention T may be called a formal semantic theory. It is a theory about the contributions of basic signs to the truth-conditions of sentences. As shown by Davidson (1967), a theory like this can be developed as an empirical theory of natural languages. We can formulate hypotheses about what the semantic values of the basic signs of the language are, and then test them by checking which sentences under which conditions the speakers of the language are willing to assent to. When the theory is developed for logical purposes, however, which is where our focus is, we are interested instead in what relations between sentences are reflected in the theory *regardless* of the values of the basic signs.

Given Convention T, a pairing of sentences with truth-conditions must be sufficient for an understanding of what truth is. It must consequently also be sufficient for an understanding of the laws of truth. What is needed to explain the nature logic, as Peter Sullivan puts it, is thus “an account which presents the ways in which propositions relate inferentially to one another as but another aspect of how they severally relate to the world they describe” (2001, p. 91). Or in other words, to understand the laws of logic as the laws of truth is to assume that the nature of logic is fully captured in a formal semantic theory. Nothing beyond what is needed to respect Convention T can be required for a recognition of why something follows logically from something else.

We have now seen what Convention T is and how it is related to logic, but why should it be accepted as a standard of adequacy for a definition of truth in the first place? What Tarski

³⁰ By this I do not mean to suggest that Anscombe does not realize that a general understanding of truth is what is responsible for excluding the division into ‘the cat is of’ and ‘ten drunk’. The quote is simply illustrative.

himself says about it is only that Convention T captures more precisely the idea behind what he calls “the classical conception of truth” (*SCT*, p. 343) that he gathers from Aristotle’s definition of truth:

[T]o say of what is that it is not, or of what is not that it is, is false, whereas to say of what is that it is, or of what is not that it is not, is true. (*Metaphysics*, 1011^b1, Γ, 25)

But it is not clear how a one-liner from Aristotle could add any legitimacy to Convention T, especially given that we are supposed to understand Aristotle’s formulation to be less precise than Convention T itself. A better explanation is given by Davidson.

For all its familiarity, I think a central point about Convention T has been largely lost on philosophers. It is this: Convention T and T-sentences provide the sole link between intuitively obvious truths about truth and formal semantics. Without Convention T, we should have no reason to believe that truth is what Tarski has shown us how to characterize. (Davidson 2001, p. 66)

As we saw previously, homophonic T-sentences are trivial. They are, that is, “intuitively obvious truths about truth”. Because of this they may be said to capture our intuitive notion of truth. And it is *this* notion—and not something invented by a theorist—that we want a precise definition of. A definition that adheres to Convention T thus delivers a notion that we can recognize as the notion of truth without any need for a further account to tell us that. That is why Tarski can adopt it as a criterion against which to check *any* putative definition of truth, including Aristotle’s (which Tarski finds to be quite good on that basis). A definition that does not adhere to Convention T is not recognizable as telling us about truth. For example, a theory that fails the test would be one that implied, instead of the trivial T-sentence, that ‘snow is white’ is true if and only if it coheres with some specified set of other sentences.³¹

What is it that an adequate definition captures about our intuitive notion of truth? T-sentences state the truth-conditions of object-language sentences. But there are at least two ways we can understand ‘truth-condition’. In one sense, given that T-sentences are true biconditionals, a theory that implies a T-sentence for every sentence of the object-language will get the extension of ‘true’ right—it will classify all and only the true sentences as true. This is one sense of ‘truth-condition’ in which T-sentences reflect the truth-conditions of sentences. They say of each object-language sentence when it is true and when it is not true.

³¹ This is a straw-man version the coherence theory of truth. Any genuine coherence theory developed after Tarski would, presumably, be developed with Convention T in mind.

But this goal can be met in many ways. For example, both of the following will serve equally well to classify ‘snow is white’ in the right way:

- (1) ‘snow is white’ is true if and only if snow is white
- (2) ‘snow is white’ is true if and only if snow is white, and $2 + 2 = 4$

Whichever one we use to classify ‘snow is white’ as true or as false, we end up classifying the sentence in the same way. But the first biconditional also tells us *what is required* for the truth of ‘snow is white’ in a sense that the second does not. It is, as Davidson emphasizes, intuitively obvious that for ‘snow is white’ to be true, it is not required that $2 + 2 = 4$. By reflecting truth-conditions in this sense, T-sentences not only provide an intuitive classification of sentences into those which are true and those which are not, but they also track our prior understanding of what it is for a given sentence *to be* true. It is because of this that a theory that adheres to Convention T captures not only the distinction between sentences that are true and sentences that are not but also the notion of truth that is relied on in drawing the distinction.

2. Tarski’s Truth by Satisfaction

Tarski’s recipe for a definition of truth has two steps. First, we define a technical notion of *satisfaction*, and then we define truth in terms of satisfaction. Tarski’s reason for introducing the notion of satisfaction is a further expectation that we have for a definition of truth. We noted that a definition of truth should apply to a language with infinitely many sentences. But we also expect the language to include generalizations, sentences like ‘someone is reading’, and ‘whoever loves someone is loved by them’. Tarski’s recipe takes for granted that generalizations have the syntax that Frege discovered them to have. Tarski sees a need for a definition of satisfaction because of how he understands this syntax.

We will look at Frege’s account of generalizations in more detail below. But to understand why Tarski takes the route that he does, we need to say something about it already. With his account of the syntax of generalizations, Dummett says, Frege “resolved, for the first time in the whole history of logic, the problem which had foiled the most penetrating minds that had given their attention to the subject” (*PoL*, p. 8). The problem was to explain the mistake made by the White King in the following exchange:

“I see nobody on the road,” said Alice.

“I only wish *I* had such eyes,” the King remarked in a fretful tone. “To be able to see Nobody! And at that distance too! Why, it’s as much as *I* can do to see real people, by this light!” (Carroll, *Through the Looking-Glass*, p. 194)

In the passage, Carroll pokes fun at the scholastic account of generality which construes expressions of generality like ‘somebody’ in terms of manners of referring.³² Obviously, it was clear before Frege that the King’s reasoning is faulty, but it was Frege who explained why it is. The King treats ‘nobody’ as a subject-term based on the similarity between ‘I see nobody on the road’ and ‘I see the messenger on the road’. Frege’s discovery, which led to our modern theory of quantification, was that this apparent parallel misleads us about the real syntactic structure of these sentences. Although the word ‘nobody’ seems to occupy the place of a subject-term, it in fact indicates a place for a subject-term *to* occupy. The word ‘nobody’ itself attaches to something that is like a sentence but has a place for a subject-term without a term occupying it. To make this apparent in the notation in a way that would work for all generalizations without ambiguity, Frege introduces the now familiar notation of quantifiers and variables. Given Frege’s account, we can paraphrase Alice’s statement as:

‘It is not the case that there is x such that I see x on the road’

The possibility of reading this as concerning a thing under the guise of Nobody has thereby disappeared.

Just like the scholastic logicians did with natural language, Tarski takes the surface-grammar of the Fregean notation at face-value. He consequently sees quantified sentences as decomposing into a quantifier and an expression like ‘I see x on the road’. This creates a difficulty. Since the definition of truth should proceed recursively, it should characterize the truth-conditions of generalizations partly in terms of the semantic values of such expressions. But these expressions may in turn be complex and their values would thus in turn have to be determined by the values of other expressions. To account for this, Tarski introduces what he calls sentential functions—known as formulas in contemporary terminology—which are like sentences but “may contain the so-called free variables (like ‘ x ’ and ‘ y ’ in “ x is greater than y ”), which cannot occur in sentences” (*SCT*, p. 352). Tarski’s definition of truth then proceeds by a recursive definition of satisfaction for formulas which is subject to a constraint similar to Convention T.

³² See Dummett, *Pol*, Ch. 2 for a proper overview of the problem that Frege solved.

Satisfaction, as Tarski introduces the notion, is a relation between sequences of objects and formulas. Each slot in a sequence of objects is taken to correspond to a specific variable. Following contemporary terminology, we may call these sequences variable-assignments.³³ The constraint for the recursive definition is that it should yield sentences of the form³⁴

(S) for every variable-assignment g , g satisfies the formula ϕ if and only if p

In (S), ' p ' is the result of taking the formula ϕ or its translation and replacing its free variables with terms that refer to the objects corresponding to the variables in the variable-assignment g . If the first slot in a variable-assignment corresponds to the variable ' x ', we should get, for example, ' $\langle a, b, \dots \rangle$ satisfies ' x is white' if and only if a is white'.

For a recursive definition of satisfaction, we also need to classify formulas according to their compositional structure. In a formal language, this can be achieved by defining formulas themselves recursively: we first give rules that specify what the simple formulas are like, and then give additional rules which characterize complex formulas in terms of other formulas. The standard language of first-order logic is an illustration of this. We then define satisfaction by first stating the conditions under which a given sequence of objects satisfies a simple formula, and then introducing rules that specify how the satisfaction of a complex formula depends on the satisfaction of formulas in terms of which it is characterized. The result is a definition of satisfaction which contains one clause for each type of formula, and from which the satisfaction-condition of any given formula can be derived.

The second step in Tarski's recipe is to define truth in terms of satisfaction. In T-sentences, truth is not attributed to formulas with free variables (variables not bound by any quantifier). It is attributed to sentences only. To account for this, Tarski defines a sentence as a formula that contains no free variables. Since the definition of satisfaction applies to all formulas, it also applies to sentences. This leads to Tarski's proposed definition.

It turns out that for a sentence only two cases are possible: a sentence is either satisfied by all objects, or by no objects [that is, by variable-assignments]. Hence we arrive at a

³³ Equating the contemporary notion of a variable-assignment with Tarski's notion of a sequence of objects may be slightly misleading. Variable-assignments are sometimes thought of as assignments of *referents* to variables. Reference is a semantic notion. In Tarski's definition of truth, no semantic relation is assumed between variables and objects. All that is assumed is that positions in a sequence of object map onto variables. But we may also treat the contemporary notion of a variable-assignment as involving no more than that.

³⁴ See Tarski 1956, pp. 191–192 for a more precise formulation.

definition of truth and falsehood simply by saying that *a sentence is true if it is satisfied by all objects, and false otherwise.* (SCT, p. 353)

So, the definition of truth we arrive at is the following:

A sentence ϕ is true iff every variable-assignment satisfies ϕ .

This completes Tarski's definition of truth. It is the definition that one finds in most contemporary logic textbooks. Although we only gave an outline of it, it should be clear that it does satisfy Convention T. Every formula receives a satisfaction-condition in the compositional definition of satisfaction. This condition will amount to a truth-condition when the formula is a sentence. The definition suffers from a different problem, however.

If a definition of satisfaction has clauses for logical operators, then there will be formulas that contain free variables, like ' $Fx \vee \neg Fx$ ', that are satisfied by all variable-assignments. And there will likewise be formulas containing free variables, like ' $Fx \wedge \neg Fx$ ', that are satisfied by no variable-assignments. The reason why they do not count, respectively, as true and as false is only because they are not sentences. This means we cannot read Tarski's proposed definition of truth as a biconditional. We must instead read it as follows:

If ϕ is a sentence, then ϕ is true if and only if every variable-assignment satisfies ϕ .

And a claim of this form is not suitable for being a definition of truth. By it, we say when a *sentence* is true, but we do not exclude the possibility of the same notion of truth applying to other things besides sentences. And we have said nothing about what is required for the truth of those other things, whatever they may be.

Let us put the objection more carefully. The problem is not the mere fact that Tarski's proposed definition takes the form of a conditional, but that the antecedent of the conditional is not redundant. For example, if we ignored Convention T, the following conditional could be a legitimate definition:

If ϕ is a formula, then ϕ is true if and only if every variable-assignment satisfies ϕ .

The exemplars of 'every variable-assignment satisfies ϕ ' are defined in our semantics for formulas only. There is thus nothing else that ϕ could be, other than a formula, so that the whole would be meaningful.³⁵ To replace ' ϕ ' with 'a shoe', for example, does not give us a *false* instance because we have specified no meaning to 'every variable-assignment satisfies a

³⁵ Since the exemplars of ' ϕ is true' are being defined, ' ϕ is true' adds no constraint on possible values of ' ϕ '.

shoe'. It gives us nonsense. So, the antecedent of the whole conditional, though we may include it to improve intelligibility, is redundant. It excludes nothing from the range of the definition. Because the antecedent in Tarski's proposed definition is not redundant in this way, the proposed definition tells us what truth amounts to in the case of sentences but does not exclude the possibility of it amounting to something else in other cases. For example, we could leave everything in Tarski's definition in place and add:

If ϕ is not a sentence, then ϕ is true if and only if *some* variable-assignment satisfies ϕ .

This would be consistent with everything introduced into the definition previously. And thus, Tarski's definition cannot be telling us what truth *is*.

The problem leaves us faced with a dilemma.³⁶ One way to get over the immediate issue would be to allow truth to be applied to any formula, not just to sentences. This would let us see the definition of truth given in Tarski's second step as a bi-conditional. However, a definition like this would not conform to Convention T. A formula '*x* is reading' is satisfied by each variable-assignment just in case each object in the domain is reading. Thus, among the implications of a definition of truth now envisaged would be

'*x* is reading' is true if and only if everyone is reading.

Biconditionals like this are not intuitively obvious truths about truth. Not only do we no longer have a direct mirroring between the expression mentioned on the left-hand side and the expression used on the right-hand side, but the expression mentioned is not something about which there *could* be intuitively obvious truths. Formulas with free variables are introduced in a Tarskian definition as theoretical entities to make sense of generality. There cannot be truths that concern them by which the adequacy of *any* definition of truth could be assessed.

Alternatively, we might dispense with the goal of defining truth altogether and instead treat the definition of satisfaction itself as central to logic, thereby doing away with the problematic second step in Tarski's recipe altogether. Since the resulting account would not aim to be a definition of truth, it would not be appropriate to expect it to conform to Convention T. However, the sole reason for defining satisfaction was to get around a technical difficulty in providing a definition of truth. Once truth is taken out of the picture, there is no reason left to keep satisfaction in the picture, either. Furthermore, our goal was not to have some account or other of logic but to have an account which construes it in terms of the laws of *truth*. By giving

³⁶ I am grateful to my examiner Peter Milne for pushing me to clarify the problem.

up the idea of truth playing a central role in logic we also lose sight of any special relationship there may be between logic and *knowledge*, thus defeating the point of exploring the nature of logic in the context of apriority.

The source of the problem with Tarski's definition is the attempt to define truth in terms of satisfaction. It is noteworthy that Tarski himself does not seem entirely happy about the idea of defining truth in that way.

It may seem strange that we have chosen a roundabout way of defining the truth of a sentence, instead of trying to apply, for instance, a direct recursive procedure. The reason is that compound sentences are constructed from simpler sentential functions, but not always from simpler sentences; hence no general recursive method is known which applies specifically to sentences. (*SCT*, p. 353)

What Tarski notes is that, had it not been for generalizations, we could have defined truth itself recursively without a detour through a definition of satisfaction. We could have first specified what it is for simple sentences to be true and then introduced rules which determine the truth-conditions of all other sentences from the truth-conditions of the simple ones. But alas, he adds, "no general recursive method is known" by which the truth-conditions of generalizations could be defined in terms of truth-conditions of other, simpler sentences. But a closer look at Frege's account reveals that such a method was known after all.

3. Thoughts

Frege never gave a definition of truth of the sort Tarski had in mind. But we can give one based on Frege's ideas. The Fregean definition, though like Tarski's in many ways, is a direct recursive definition of the sort Tarski thinks impossible. And it is this definition that we will use to understand logic as the study of the laws of truth. To develop the definition, we must first introduce Frege's notion of a thought.

What is distinctive about my conception of logic is that I begin by giving pride of place to the content of the word 'true', and then immediately go on to introduce a thought as that to which the question 'Is it true?' is in principle applicable. (*PW*, p. 253)

The question 'Is it true?' can receive an affirmative or a negative answer. When the answer is affirmative, the thought is true. And when the answer is negative, we say that the thought is

false. So, a thought, by definition, is what can be true or false. Whatever else we say about thoughts must stem from this minimal characterization. Frege continues the above passage.

So I do not begin with concepts and put them together to form a thought or judgment; I come by the parts of a thought by analysing the thought. (ibid.)

In other words, whatever structure (if any) thoughts do have, it must be an upshot of their truth-aptness. This fits with our point from before about ‘the cat is often drunk’. But we no longer have any possibility of confusion that stems from apparent structure.

Frege’s starting point is thus significantly different from Tarski’s. Tarski *picks* sentences to be the primary truth-bearers because “[f]or several reasons it appears most convenient”, and he picks them out from a given set of candidates (*SCT*, p. 342). Frege *defines* the primary truth-bearers in terms of truth. And, *pace* Tarski, an answer to a question ‘what is true?’ is not ‘the sentence’.

[W]hen we call a sentence true, we really mean its sense is [that is, the thought expressed].
From which it follows that it is for the sense of a sentence that the question of truth arises in general. (Frege, *T*, p. 292)

Even when we ask whether the sentence ‘snow is white’ or an utterance of it is true, we want to know whether it is true *that* snow is white. Frege’s primary notion of truth is thus not expressed by the truth-predicate ‘ ϕ is true’ which attributes something to a sentence mentioned. It is expressed instead by the truth-operator ‘it is true that ϕ ’ which invokes a sentence in use.

When Frege speaks of sentences, he does not mean mere grammatical units but things which are already paired up with thoughts. But we will keep the term ‘sentence’ for mere grammatical units of the sort Tarski has in mind. We will occasionally need it. We will instead refine our previous notion of a *proposition*.³⁷ So far, we have been speaking of propositions as sentences individuated in terms of their truth-conditions, rather than their grammatical structure. The notion of a thought lets us define a proposition simply as an *expression* of a thought. (As we will see, the two notions of proposition amount to the same.) Propositions, thus understood, will make up our object-language. And like before, we will assume that the

³⁷ Frege’s term that is often translated as ‘sentence’ is ‘Satz’. In J. L. Austin’s translation of *Foundations of Arithmetic*, it is translated as ‘proposition’. We will follow Austin in this.

object-language has infinitely many propositions built out of a finite vocabulary (though the constraints on vocabulary will be understood slightly differently).

We will adopt a truth-predicate for attributing truth or falsity to proposition. But, in line with Frege's point from before, we will always understand this in terms of the truth or falsity of the thought that the proposition is an expression of. Since propositions are expressions of thoughts, there is, by definition, a thought for each proposition. Given finite vocabulary, which thought is expressed by a proposition must be determined compositionally by the parts of the proposition, with each part making its distinctive contribution to the determination of a thought. Consequently, a thought too must have parts and a compositional structure.

The idea of including thoughts in Frege's sense into a semantic theory is not new. Contemporary semantic theories which are developed with philosophical aims in mind often try to accommodate them. In a semantic theory, we have for each sentence '*p*' its T-sentence: '*p*' is true iff *p*. The standard approach, what Recanati (2012) has called a two-level semantics, is to *add* thoughts to a semantic theory by re-formulating the T-schema like so:

(T') the thought expressed by '*p*' is true iff *p*

In two-level semantic, sentences are first paired up with thoughts, and then a proper formal semantics is developed for thoughts, which will give us T'-sentences.

The two-level approach is designed to allow for the possibility of two propositions expressing different thoughts with the same truth-condition. The distinction in thought is then invoked to explain the possibility of propositional attitude reports like 'S believes that Hesperus is bright' and 'S believes that Phosphorous is bright' having different truth-values. And the sameness of truth-condition of those thoughts is invoked to account for the fact that the same heavenly body, the planet Venus, is mentioned in both reports. The problem of accounting for both features of propositional attitude reports is known as Frege's puzzle, and the two-level semantics solution to it is often seen as Frege's solution.³⁸

The attribution of the idea of a two-level semantics to Frege is natural if we take formal semantics as a paradigm for understanding truth. A thought, since it has a truth-value, is assumed to need a truth-condition. Coupled with the idea that sentences express thoughts, it simply follows that Frege maintained a two-level semantics. An upshot of a two-level

³⁸ For a detailed proposal of two-level semantics see also Perry 2001, in addition to Recanati 2012. The text commonly referenced as the source of the puzzle is Frege's "On Sense and Reference".

semantics is that the parts of a thought, like the parts of a sentence in standard semantics, must contribute systematically to the determination of truth-conditions. The result is an account where thoughts are construed as having many features that we normally take sentences to have. And on some versions of two-level semantics, like the one given by Fodor (2008), thoughts are simply taken to be sentences of a special sort.

Whatever the merits of a two-level semantics, it misconstrues Frege's notion of a thought. Frege maintained that the truth of a thought is indefinable, as we will see. But in T-sentences, we define the truth of a sentence in terms of the obtaining of a truth-condition. And in T'-sentences of a two-level semantics, we do the same for the truth of a thought. Thus, the two-level semantics that is sometimes attributed to Frege cannot be Frege's. However, it is perhaps easier to see this as a problem for Frege than for two-level semantics. Surely, it might be said, truth *must* go along with truth-conditions.

In response to the objection, the first thing to note is that a formal semantic theory is a definition of truth for sentences. Although we replaced sentences with propositions, we have already granted that truth can be defined that way. But the objection could be elaborated: surely, a definition that can be given for sentences or propositions can also be given for whatever else truth is attributed to, including thoughts. This is indeed what Tarski himself maintains, noting that defining truth for sentences "does not exclude the possibility of a subsequent extension of this notion" (*SCT*, p. 342). And it is here that Frege gives us a simple but powerful argument that this is not so.

[I]n a definition, certain characteristics would have to be stated. And in application to any particular case the question would arise whether it were true that the characteristics were present. So one goes round in a circle. Consequently, it is probable that the content of the word "true" is unique and indefinable. (*T*, p. 291)

We cannot, according to Frege, define truth, though we can rely on the indefinable notion to extend the use of 'true' beyond its primary application to thoughts. But the argument calls for some clarification.

Frege's point is not that a definition of truth itself is inevitably circular. The circularity ensues, rather, in one's attempt to recognize something as falling under the definition. To check whether the definition applies in a given case, one needs to check whether its definiens is true in that case. But an adequate definition should be useful. Someone who is unable to tell whether a given thing is *F* without a definition of *F* should be able to do so by relying on the definition.

If one is unable to tell without a definition of *truth* whether a thought is true, then, like with all definitions, relying on the definition in a given case will give rise to the question whether the definiens is true in that case. Since, as stipulated, the person is unable to tell without a definition of truth whether a thought is true, they must now resort to the definition again. And so, we have a circularity.

Dummett (*Pol*) presents two objections to Frege's argument. First, he notes that a similar reasoning seems to show that "the notion of truth had to be rejected altogether, whether defined or not" (p. 443). As Dummett reads the argument, Frege assumes that to check whether *p*, we should first check whether it is true that *p*, and thus whether it is true that it is true that *p*, etc. And so, it seems that "[t]he possibility of the regress . . . has nothing to do with whether truth is definable or not" (ibid.). Dummett's second objection is that, even if there was a regress, it would be vicious "only if it were supposed that, in order to determine the truth of any member of the series, I had first to determine that of the next term in the series" (ibid.). But, Dummett says, we may suppose that the truth-values of all propositions in the series are determined simultaneously.

Dummett gets Frege's argument the wrong way around. Both of his objections amount to the same worry, that we need not put the question 'is the definiens true?' in terms of truth. The reason is that to ask whether the definiens is true is the same as to ask whether *p*, where '*p*' is the definiens. But this is precisely Frege's point.³⁹ Since we ask the *same thing* by both questions, it does not matter which formulation we choose. To fail to know whether it is true that *p* is to fail to know whether *p*. One therefore will not be able to tell whether *p* without a prior understanding of truth. The assumption here is that the truth-operator 'it is true that ϕ ' contributes nothing to the thought. But this assumption we have already defended. As we noted in the previous chapter, if the truth-operator made a further contribution to a thought, then '*p*' and 'it is true that *p*' would express different thoughts. Even if we supposed that there is some common metaphysical underpinning for them, a subject would still have to establish the truth of the second in addition to the truth of the first. An investigation into whether *p* would thus be distinct from an investigation into whether it is true that *p* and truth would not be construed as the aim of all investigation. Since this cannot be, the thought that *p* must be the very same

³⁹ This is clear from Frege's earlier formulation of the argument. When considering the definition of truth in terms of correspondence to reality, Frege notes that to apply the definition "we would have to ask in each case whether an idea corresponds to reality, *in other words*, whether it is true that the idea corresponds to reality" (1969, p. 140; cited from a translation in Sluga 2002; emphasis added).

thought as the thought that it is true that p , regardless of our choice of ‘ p ’. Frege’s argument for the indefinability of truth now runs.

We can now return to merging thoughts into a formal semantic theory. Since we can freely switch between ‘ p ’ and ‘it is true that p ’, we can also paraphrase a T-sentence—a sentence of the form ‘‘ p is true iff q ’—as ‘‘ p is true iff it is true that q ’, and *vice versa*. The thought, therefore, *just is* the truth-condition. For a thought to be true is for that thought to be *so*. As Frege himself puts it, “[a] fact is a thought that is true” (*T*, p. 307). This makes the T-sentence of a proposition a statement about which thought is expressed by the proposition.⁴⁰ It is a definition of the truth of the proposition in terms of the truth of a given thought. And a general definition of truth for all propositions of the object-language is thus a definition of their truth in terms of the truth of thoughts. Since we are assuming that the object-language has infinitely many propositions, this definition should proceed recursively, just like Tarski’s. As Tarski’s attempt illustrated, an obstacle in constructing a definition like this arises with the attempt to accommodate generality. Tarski adopted from Frege only the notation of quantifiers and variables. This is what led to the problematic separation of truth and satisfaction. To avoid it, we need to take a closer look at Frege’s account of generality.

4. Frege on Generality

As we saw, Frege’s insight about the syntactic structure of generalizations like ‘everything is red’ was that ‘everything’ in the sentence indicates a place where a subject-term goes, rather than occupying that place itself. The phrase should instead be thought of as a quantifier attaching to an expression that is like a sentence but has no subject-term in the place indicated by the quantifier. Frege’s paraphrase of the sentence would be ‘every x is such that x is red’, or in pseudo-formal language, ‘ $\forall x x$ is red’. Tarski parsed a sentence like this into a quantifier and the formula ‘ x is red’, which he then included among the simple formulas that admit of no further analysis. But for Frege’s account of generalization to work, the remainder of a general proposition, once the quantifier is removed, is an expression that *must* admit of further analysis.

⁴⁰ Although we will later see that this is not quite right.

Let us call sentences like ‘*a* is red’, ‘*b* is red’, etc. the *instances* of the generalization ‘ $\forall x$ *x* is red’. We may think of Frege’s account of generality as aiming to answer this question: how is the truth-value of a generalization linked to the truth-values of its instances? If one of the instances is false, the generalization ‘ $\forall x$ *x* is red’ is also false, and if all instances are true, the generalization is also true. The problem is that the thoughts expressed by the instances cannot themselves be given by the thought expressed by the generalization. As Frege puts the point: “We must not think that I mean to assert something about an African chieftain . . . who is wholly unknown to me, when I say ‘All men are mortal’” (*PW*, p. 213). If, for each individual person, we spoke of *them* when uttering ‘all men are mortal’, then, to recognize the thought expressed by the generalization, we would have to grasp one thought for each individual in the domain of the generalization. And, as Dummett (*PoL*) emphasizes, this is not possible for generalizations over infinite domains, like ‘between every two points there is a line’. That is why a generalization cannot, for example, be thought of as a compacted conjunction of its instances.

To explain the link between the truth-values of the generalization and the truth-values of its instances, Frege introduces the idea of incomplete (or unsaturated) expressions.

Statements in general . . . can be imagined to be split up into two parts; one complete in itself, and the other in need of supplementation, or unsaturated. Thus, e.g., we split up the sentence ‘Caesar conquered Gaul’ into ‘Caesar’ and ‘conquered Gaul’. The second part is unsaturated—it contains an empty place; only when this place is filled up with a proper name, or with an expression that replaces a proper name, does a complete sense appear. (Frege 1960, p. 31)

We get an incomplete expression by “splitting up” a proposition. The reason why “splitting” a proposition yields something more than just a shorter typographical string is that a proposition expresses a thought (“a complete sense”), that is, it has a truth-condition. The “splitting” of a proposition, as emphasized by Dummett (*PoL*, pp. 30–31), is to be understood not as a decomposition of a syntactic unit into its parts but as an *operation* on a proposition. It is definitive of the expression that it is the result of such an operation, that it is an expression “formed by omission of a proper name” from a proposition (*ibid.*, p. 23). The expression operated on—the proposition—expresses a thought but the result of the operation, the incomplete expression, does not. Yet, since the incomplete expression is to be understood as the result of omitting a name from a proposition, it is still the sort of expression whose completion with a name yields a thought.

As Dummett explains Frege’s idea, an incomplete expression should be “regarded as a *feature* in common” between different propositions (*PoL*, p. 31). Like all propositions, the propositions ‘*a* is red’, ‘*b* is red’, ‘*c* is red’, etc. express thoughts and thus have truth-values. What separates these propositions from others is that there is also something in common between them. They are all of the same form, namely they are all propositions where a name occurs to the left of ‘is red’. It is by means of the shared form that we can, as Ramsey (1925, p. 124) puts it, “collect together” the propositions. Frege’s insight is that this shared *form* can itself be seen as an expression with its own characteristic semantic value.

This kind of appeal to common features of expressions is still recognized as essential in contemporary logic (and was used also by Tarski). Without it, one cannot state general rules pertaining to sentences depending on their form alone. In contemporary notation, the so-called Quine-quotes along with meta-variables are often adopted for this purpose. For readability, I will always use Greek letters for meta-variables and will reserve ‘ ϕ ’, ‘ ψ ’, and ‘ χ ’ for sentence-variables. For example, ‘an expression that consists of ‘ \neg ’ followed by a sentence ϕ ’ is abbreviated with the help of Quine-quotes as: ‘ $\neg\phi$ ’. The incomplete expression understood as a feature in common between ‘*a* is red’, ‘*b* is red’, etc. will thus be written as: ‘ ξ is red’, where ‘ ξ ’ is a meta-variable indicating the place where a name occurs.

Since the incomplete expression ‘ ξ is red’ is to be understood as a common form of ‘*a* is red’, ‘*b* is red’, ‘*c* is red’, etc. it cannot be understood as a *part* within each one. As Dummett explains it,

because the ‘complex’ predicate [the incomplete expression] ‘ ξ snores’ has to be regarded as formed from such a sentence [proposition] as ‘Herbert snores’, it cannot itself be one of the ingredients from which ‘Herbert snores’ was formed, and thus cannot be that whose sense . . . contributes to composing the sense of ‘Herbert snores’. (*PoL*, pp. 30–31)

It is only because each proposition of the form ‘ ξ is red’ expresses a thought that we can understand ‘ ξ is red’ itself as an incomplete expression and its empty place as calling for a name. The expression ‘ ξ is red’ is a form whose exemplars are propositions ‘*a* is red’, ‘*b* is red’, ‘*c* is red’, etc. We get from the form to the exemplar by choosing a name. There are, of course, other expression that we could choose to get propositions. For example, we could choose ‘nothing’ to form ‘nothing is red’. But this will not be a way of completing the incomplete expression because the resulting proposition is not an exemplar of ‘ ξ is red’.

If the incomplete expression is to be understood as a common form of different propositions, what should be its semantic value? Let us assume some truth-values for each exemplar of 'ξ is red', so that, for example, 'a is red' is true, and 'b is red' is false. A characterization of the semantic value of 'ξ is red' would then be the following: if its empty place is filled with a name of *a*, then the result is a true proposition, if its empty place is filled with a name of *b*, then the result is a false proposition, and so on. We can write it as a table

<i>a</i>	↔	true
<i>b</i>	↔	false
⋮	⋮	⋮

that records how the referents of names are paired up with truth-values. What we have, in other words, is a function from objects—the referents of names in the exemplars—to the truth-values of the exemplars. Frege calls such functions *concepts*—"a concept is a function whose value is always a truth-value" (1960, p. 30)—and the expressions whose semantic values they are *concept-expressions*.

We have seen why concept-expressions are incomplete, but Frege also calls concepts themselves incomplete.

[T]he argument does not belong with the function, but goes together with the function to make up a complete whole; for the function by itself must be called incomplete . . . (Frege 1960, p. 24).

It is at any rate not obvious in what sense a function—a pairing of objects to truth-values—could be incomplete. Yet, Frege does not go to great lengths to distinguish between the incompleteness of the concept-expression and the incompleteness of the concept at all. The reason is that one is an immediate consequence of the other. A concept expressed by 'ξ is red' is a function from the referents of the names in the exemplars to the truth-values of the exemplars. But to tell which exemplars the concept-expression has and thus the truth-values of which thoughts are the values of the concept, we need to know *which* names can be used to complete the concept-expression. This cannot be read off the concept-expression, since the concept-expression is precisely the result of letting the names go variable. But if we let the names go variable, we also let the *referents* of those names go variable. The concept-expression 'ξ is red' is the form of its exemplars, and the concept expressed by it is the form of the thoughts of its exemplars. We thus cannot represent the concept as a function-table when considering just the concept-expression itself. It doesn't have the information encoded in it that is needed to produce the table. We can only characterize it in general terms:

' ξ is red' is true if and only if the referent of the name ξ is red.

Unlike an explicit list, this characterization includes a meta-variable indicating the place where a name would go. But by leaving a place for a name, it also leaves a place for the referent of a name in the right-hand side. This characterization doesn't tell us which objects are the arguments of the function any more than it tells us which names occur in the exemplars of ' ξ is red'. Still, since the expression does have exemplars which do express thoughts, each of which is either true or false, the general characterization specifies a function.

The result we have arrived at is familiar from contemporary semantics, but it is not the same. In their influential textbook, Heim & Kratzer (1998) give an illustration of how to "execute the Fregean program" (p. 13). They use 'Ann smokes' as an illustration. Taking inspiration from Frege's splitting-operation, they divide 'Ann smokes' into two parts: 'Ann' and 'smokes', and then assign a concept to 'smokes'. We have seen what this construal gets wrong. A part of a proposition does not have a concept as its value. Only a common form of propositions does. This mistake, though seemingly minor, then gives rise to a problem at the heart of their semantic project. They realize that, given how they have set things up, there is a choice in how to present the concept of 'smokes'. It could be presented as a function-table or by stating a condition as we have done for the concept of ' ξ is red'. But, they admit, only the second mode of presentation could yield a truth-condition and thus reflect the meaning of the expression. A function-table, they say, is "certainly not what we have to know in order to know the meaning of "smoke"" (p. 21). So, they try to motivate the choice in favour of presenting concepts by conditions.

We didn't really have a choice, though, because displaying the function in a table would have required more world knowledge than we happen to have. We do not know of every existing individual whether or not (s)he smokes. (ibid.)

What they point to, namely that we do not know of each individual whether they smoke, is just the problem we had to avoid in constructing a semantics for generalizations. But this is not an *explanation* of why the concept must be presented by a condition. It is a recognition of a need for an explanation. But Heim & Kratzer have no way to give such an explanation since the reason why the concept can only be presented by a condition is precisely that it is a form of propositions in which a name is unspecified. Only then is the presentation of a concept as a function-table excluded.

We have now covered Frege's concept-expressions and their semantic values. Based on this, we can provide a semantics for generalizations which does not require a definition of satisfaction. We can specify the syntax of quantified propositions as follows. If ϕ is a proposition in which a name α occurs, then replacing one or more occurrences of α with a variable ν and attaching '∀ ν ' in front of the result is a proposition. Or simply: if $\phi(\alpha)$ is a proposition, then '∀ $\nu\phi(\nu)$ ' is a proposition. A concept-expression is invoked in the antecedent of this specification. By selecting the initial proposition $\phi(\alpha)$ as a proposition in which a name occurs, we specify a place for a name. There is no need for anything like empty slots in the object-language. When we attach a quantifier to the concept-expression, we replace whatever name occurs in its name-place with a variable bound by the quantifier. Unlike in the standard Tarski-inspired notation, we will never need free variables. The semantic axiom for universal quantification can now be given as follows:

'∀ $\nu\phi(\nu)$ ' is true iff for every α' , $\phi(\alpha)$ is true

where $\phi(\alpha)$ is the result of removing the quantifier and replacing the variable ν with a name α in all its occurrences where it is bound by the quantifier, and α' is the referent of α .

The resulting account bears some resemblance to what are known as substitutional accounts of generality, accounts where the instances of a generalization are defined with the help of a domain of names. A common objection against this way of specifying a semantics for generalizations is that it presupposes that we have a name for every object. A generalization should have an instance for every object. But in a substitutional account, we have only as many instances of the generalization—only as many propositions of the form $\phi(\alpha)$ —as we have names. So, every object should have a name in the language. The objection is that each of those names would somehow need to be specified. But there are limits to how many names we can specify for the language. If we introduced the names one-by-one, we would only have a finite number of names and, given a substitutional account of generalizations, we could only generalize over finitely many things. But even when we specify names recursively by a rule for constructing new names from existing names, we would still have only denumerably many names and could thus not generalize over real numbers or points on a plane, for example.

Dummett explains why this objection, though it may apply to substitutional accounts, does not apply to Frege's account. Unlike substitutional accounts, Frege's account does not presuppose that we already *have* a name in the language for every object.

Frege's assumption is that, if we understand 'A(c)', then we likewise understand, for any object whatever, what the truth-conditions would be of a sentence formed by putting a name of that object, rather than the name 'c', in the argument-place of 'A(ξ)'. (*PoL*, p. 17)

In other words, what is assumed is that we have a general understanding of what a name *does*. We need not have a specific range of names in view, so to speak. We already saw why this is so in our explanation of what a concept-expression is. The expression ' ξ is red' is the common form of all propositions where a name is followed by 'is red'. We saw that this expression can be extracted from a single proposition, like '*a* is red'. We cannot gather from '*a* is red' alone which names there are in the language. All that we can recognize—and all that we *need* to recognize—is what sort of role is played by a name in a proposition. This is enough for the possibility of always *introducing* a name when we need one by adopting a sign for it.

In contemporary formal logic, it is customary to think of languages as being partly individuated by which names occur in them. If we stick to this conception of language, then Frege's account requires an open-ended range of languages that all differ from each other with respect to names. Alternatively, we should not think of a language as being partly individuated in terms of names at all. In specifying a language in the Fregean sense, we specify a role for names to play but do not specify which typographical signs play this role.

The role of a name, of course, cannot be understood purely syntactically. It is not enough if we think of a name as a typographical sign that may be combined in certain ways with other signs. A name is an expression that contributes an object to a thought. Here, we see again why Frege's account of generality requires the notion of a proposition rather than that of a mere sentence. To have the open-ended view of names that the account requires, we need to think of names as constituents of expressions of thought. It is also a further respect in which Frege's account of generality differs from substitutional accounts. The promise of a substitutional account is a semantics for generalizations that does not presuppose a notion of an object. Frege's account, on the other hand, invokes the notion of an objects in its construal of names and thus could not do away with it.

5. A Definition of Sentence

We are now ready to introduce a recursive definition of truth in a way that Tarski thinks impossible. We can directly characterize what it is for simple propositions to be true and then

introduce rules by which the truth-conditions of complex propositions are determined by the truth-conditions other propositions.

To avoid any potential confusions later, we will first define a formal language. Unlike in a Tarski-inspired language, we will have no formulas. We will instead define sentences themselves recursively. We will first define what the simple sentences are, and then introduce rules which characterize complex sentences in terms of other sentences. Following standard terminology, we will call our simple sentences *atomic* sentences. We will stick to the following notation for atomic sentences:

Individual signs: lower-case letters (other than ‘ x ’, ‘ y ’, ‘ z ’) with or without subscripts

Relational signs: upper-case letters (other than ‘ X ’, ‘ Y ’, ‘ Z ’) with or without subscripts

We will call both individual and relational signs simple signs. For the rules for the construction of new sentences out of existing ones, we will also introduce the standard signs for the logical operations, brackets for distinguishing scope, variables (‘ x ’, ‘ y ’, ‘ z ’ and ‘ X ’, ‘ Y ’, ‘ Z ’ with or without subscripts for replacing, respectively, individual and relational signs) and the signs for the two quantifiers.⁴¹

Definition of sentence

1. If Π is a relational sign, and $\alpha_1, \dots, \alpha_n$ are individual signs, then ‘ $\Pi\alpha_1 \dots \alpha_n$ ’ is a sentence.
2. If ϕ and ψ are sentences, then ‘ $\neg\phi$ ’, ‘ $(\phi \wedge \psi)$ ’, ‘ $(\phi \vee \psi)$ ’, ‘ $(\phi \rightarrow \psi)$ ’, and ‘ $(\phi \leftrightarrow \psi)$ ’ are sentences.
3. If $\phi(\alpha)$ is a sentence in which a simple sign α occurs and ν is a variable, then ‘ $\forall\nu\phi(\nu)$ ’ and ‘ $\exists\nu\phi(\nu)$ ’ are sentences (where α in some of its occurrences is replaced with ν).
4. Nothing else is a sentence.

For a proper understanding of the definition, we need to clarify a potential confusion about Quine-quotes. As we noted, they abbreviate structural descriptions. But it may be tempting to read the following as equivalent ways of spelling out the Quine-quotes: If Π is a relational sign and α an individual sign, then

- 1) Π ’s *standing to the left of* α is a sentence.

⁴¹ To avoid excessive use of brackets, we will adopt the standard conventions for dropping them from sentences. The signs that come with brackets are ordered according to scope-priority as follows: ‘ \wedge ’ > ‘ \vee ’ > ‘ \rightarrow ’ > ‘ \leftrightarrow ’.

2) *the result of putting* Π to the left of α is a sentence.

But they are not equivalent. In fact, the second *cannot* be a way of spelling out the Quine-quotes. It would not give us a construction of sentences but a *recipe* for construction. It specifies what a sentence is in terms of what *we* must do to *get* a sentence. A sentence is then defined as the upshot of following the recipe. If our whole definition of sentence is read in this way, it would no more give us an infinite supply of sentences than a hummus recipe gives us an infinite supply of hummus. A way to adjust the second reading of Quine-quotes would be to take it as a characterization of an operation on expressions, much like ' $n + 1$ ' may be read as 'the result of adding 1 to n ' but really means the same as 'the sum of n and 1'. But even if we take the talk of "putting" an expression next to another as a characterization of an operation, it cannot be maintained for our definition of sentence generally. There is no reason why the result of an operations should have the operands as parts. If ' Fa ' was characterized as the result of an operation with ' F ' and ' a ' as operands, it would not have ' F ' and ' a ' as parts any more than the sum of 1 and 2 has 1 and 2 as parts. But sentences *must* have parts. They must contain terms that contribute objects that make up the thought expressed.

A direct impact that reading the Quine-quotes as abbreviations of a recipe would have to the rest of our definition of truth is that we would then end up with an account of sentences where sentences themselves are construed as *objects* and thus as potential referents of object-language terms.⁴² In addition to the problems already mentioned, this would have a further undesirable consequence. To avoid paradoxes like the liar paradox, we would then need to state explicitly that the object-language sentences themselves are not potential referents of the object-language terms. That is, we would need to distinguish two kinds of objects in the meta-language: sentences and the rest. This would be a way of banning statements of the paradox but not a way of solving the paradox.

There may well be some independent theoretical value in treating sentences as objects for certain purposes. What we have said so far, does not exclude this. But sentences cannot be treated as objects in a definition of truth. A definition of truth must concern all sentences of the language, whether they are constructed or not. To ensure this, we must define sentences directly

⁴² Despite specifying sentences by description throughout his work, Tarski fails to take this point into account and thus does end up treating sentences as objects. It is notable that his illustration of how sentences can be specified by description begins with 'the sentence constituted by ...' (see *SCT*, p. 344). This sort of specification, though it might be read as identifying an object ('*the* sentence'), is not available to us if we are still in the process of defining what sentences are. We have at that point no way to understand 'sentence'.

as ways for signs to be arranged, not as the results of some specified process. This is our justification for the view we find in Wittgenstein's *Tractatus*.

Only facts can express a sense, a set of names cannot. (*TLP*, 3.142)

Instead of, 'The complex sign " aRb " says that a stands to b in the relation R ', we ought to put, 'That " a " stands to " b " in a certain relation says that aRb .' (*TLP*, 3.1432)

In other words, the compositional structure that an understanding of truth requires us to see in sentences is of the sort that is itself to be characterized by means of sentences. We will see in the next chapter how impactful this point is.

6. Atomic Propositions

Before we provide a definition of truth, some clarification is in order concerning atomic propositions, atomic sentences to which thoughts are assigned. Atomic propositions appear in Tarski's recipe for a definition of truth. Frege does not explicitly discuss atomic propositions. Wittgenstein who builds on Frege's work does and, like Tarski, recognizes their importance.

Wittgenstein construes atomic propositions, or elementary propositions, as he came to call them in the *Tractatus*, as concatenations of terms.

An elementary proposition consists of names. It is a nexus, a concatenation, of names.
(*TLP*, 4.22)

The terms themselves are construed as simple signs whose meaning is exhausted by the objects they stand for: "A name means an object" (*TLP*, 3.203). Thus, *all* simple signs are treated as receiving the same kind of semantic value: "Relations and properties, etc. are *objects* too" (*NB*, p. 61). Wittgenstein calls the thought expressed by an atomic proposition a *state of affairs*. An atomic proposition, as he puts it, "asserts the existence of a state of affairs" (*TLP*, 4.21). Just as an atomic proposition is parsed into names, a state of affairs can be parsed into objects: "A state of affairs (a state of things) is a combination of objects (things)" (*TLP*, 2.01).

Wittgenstein's construal of atomic propositions might seem foreign to a contemporary reader. Given the standard textbook account, atomic propositions are composed of a predicate that receives an extension as its semantic value and some number of terms which receive objects as their semantic values. The truth of an atomic proposition is then defined in terms of

whether a sequence of the values of its terms is in the extension of its predicate. Before we proceed, we need to see why the textbook account is based on confusion.

First, although foreign to a contemporary reader, Wittgenstein's account of atomic propositions may, with some reservations, also be discerned in Tarski's (1956) definition of truth. In Tarski's account (see p. 190), the satisfaction of simple formulas is defined directly by disquotation or term by term translation. There are no predicates whose values are defined beforehand. In a homophonic definition, Tarski's account would give us, for example

'Fa' is true iff *Fa*

where the values of the terms in the object-language are given simply by the same terms lifted into the meta-language. There is no fundamental distinction between semantic values of two different kinds. We can thus see Tarski as agreeing with Wittgenstein on atomic propositions being concatenations of names. But even if Tarski's account agrees with Wittgenstein's, is there reason to think it is right?

The problem with the textbook account is essentially what we already noted in Heim's and Kratzer's account. To construe the relational term, as we have called it, as a predicate—as an expression that receives an extension as its value—is, in effect, to construe it as an expression of a concept which yields the value true for some objects (for those in the extension) and false for others (for those not in the extension). The truth-values that the concept yields are just the truth-values of the atomic propositions which, given the textbook account, have the predicate as a component. We thus get atomic propositions from a textbook predicate in the same way as we get exemplars from a concept-expression, by choosing terms. Previously, we noted that an account like this will not be able to explain why a proposition must have a truth-condition and not simply a truth-value in a function-table. In the context of a logic textbook, this objection might not be enough. Specifying truth-conditions is not the primary aim in logic. The main purpose of a formal semantic theory in a logic textbook is to capture the relationships between the truth-values of propositions which would let us check whether the truth of some propositions determines the truth of others.

A more general objection to the idea of construing a component of a proposition as function that yields a truth-value for the proposition is given by Ramsey in "Universals" (1925). First, we can note that *'¬Fa'* and *'Fa'* both have *'F'* as a component. If the value of *'F'* was a concept, it should directly yield a truth-value only for the latter proposition. But how could *'F'* discriminate between *'Fa'* and *'¬Fa'* in this way? As Ramsey notes, it cannot.

[C]onsider the expression ‘wise’; this we use to collect together the propositions ‘Socrates is wise’, ‘Plato is wise’, and so on, which are the values of ‘ x is wise’. But this is not the only collection we can use ‘wise’ to form . . . [W]e can use ‘wise’ to collect all those in which it occurs, including not only ones like ‘Socrates is wise’ but also ones like ‘Neither Socrates nor Plato is wise’, which are not values of ‘ x is wise’ but only of the different function ‘ ϕ wise’, where ϕ is a variable. (Ramsey 1925, p. 124)

For a component of an atomic proposition, there are at least two ranges of propositions that have it as a component, one consists of all propositions in which it occurs and the other of all and only the atomic propositions in which it occurs. To specify one range rather than the other, we need to specify a form of propositions. The mere appeal to it being a component is simply insufficient. To specify a function that is guaranteed to yield all and only the truth-values of the propositions in the atomic range, we need to select that range. But for this we need the atomic form ‘ $F\xi$ ’. And this form, as we have seen, cannot be a component of ‘ Fa ’.

What this shows is that ‘ F ’, understood as a component of an atomic proposition, cannot have a concept as its value. But it does not yet show what it does have as its value. For this, we need to recognize a symmetry between ‘ F ’ and ‘ a ’ in ‘ Fa ’. Ramsey continues:

[C]an we not do with ‘Socrates’ what we do with ‘wise’, and use it to collect a set of propositions narrower than the whole set in which it occurs? (ibid., p. 125)

The very same idea that applied to ‘wise’ as well as to our ‘ F ’ obviously applies also to ‘Socrates’ and to our ‘ a ’. To spell this out, Ramsey introduces the notion of a *quality* as something that is attributed to individuals in atomic propositions.

Supposing justice and wisdom to be qualities, ‘Socrates is wise’, ‘Socrates is just’ would belong to the narrower set and be values of a function ‘Socrates is q ’. But ‘Socrates is neither wise nor just’ would not assert a quality of ‘Socrates’ . . . and would only be a value of the function ‘ ϕ Socrates’, not of ‘Socrates is q ’. (ibid.)

In short, we can apply Frege’s splitting operation to ‘ Fa ’ in either of two ways, to yield ‘ $F\xi$ ’ which can be completed by names of individuals or to yield ‘ ζa ’ which can be completed by names of qualities. Both are forms of atomic propositions which can occur in *first-order* generalizations. A proposition ‘ $\forall X X a$ ’ generalizes over qualities of a and has ‘ Fa ’, ‘ Ga ’, etc. as instances, just as ‘ $\forall x F x$ ’ generalizes over bearers of the quality F and has ‘ Fa ’, ‘ Fb ’, etc. as instances. Both generalizations range over a domain of what in a broad sense may be called *objects* and both are captured in the same semantic axiom for quantified propositions. This

shows the significance of the third of Frege’s famous principles, “never to lose sight of the distinction between concept and object” (*FA*, p. x), and how the textbook account of atomic propositions fails to adhere to it. Ramsey’s insight about relations being objects just like individuals will be relevant to us in the next chapter where we consider the idea of logical relations. If there are logical relations, Ramsey’s point would apply to them as it does to any other relations.

The parallel with Tarski’s and Wittgenstein’s account goes further. There is something about atomic propositions in the *Tractatus* that tends to raise eyebrows. Wittgenstein says:

It is *obvious* that the analysis of propositions *must* bring us to elementary propositions which consist of names in immediate combination. (*TLP*, 4.221, emphasis added)

The claim, in other words, is that an account of propositions—a definition of truth that we will introduce in the next section—is bound to assume atomic propositions. But why is this so, and why should it be obvious? Moreover, Wittgenstein’s goes on to say later that “[c]learly we have some concept of elementary propositions quite apart from their particular logical forms” (*TLP*, 5.555), and that “we know on purely logical grounds that there must be elementary propositions” (*TLP*, 5.5562). The reason why this is so is the very same reason that Tarski gave for introducing atomic propositions. The reason Tarski introduces them is not that propositions like ‘Sally is reading’ *look* atomic and thus should be treated as such in a definition of truth. Atomic propositions must be *assumed* in a definition of truth because, to respect Convention T, we must define truth recursively. And a recursive definition must have base-clauses. And, as we noted earlier, the same constraints that a theorist has when defining truth in the meta-language are also in play in any understanding of truth that speakers of the object-language have. Convention T captures the nature of truth for both. So, we *do* “[c]learly we have some concept of elementary propositions,” as Wittgenstein says. We have it because we have a concept of truth.

7. A Definition of Truth

To proceed with a definition of truth we need to specify how to speak of the semantics of the simple signs. If α is a simple sign in the object-language, we will call α' the *referent* of α , and may say that α *refers to* α' . When a simple sign is a candidate for having a referent, we

will call it a name or a term. The referent itself, whether it is an individual or a relation, may be called an *object*.

Since, in using a name, we are referring to an object, we can replace our talk of names that can be used in the meta-language with the notion of a *domain of objects*. It consists of all objects that we can refer to in the meta-language. If the object can be referred to in the meta-language, it is included in the domain. Given this way of introducing a domain of objects, we do not need to add it explicitly into our definition of truth. But we should keep in mind that we do have such a notion in the background. Given the Fregean account of generalizations, we can always introduce new names into the object-language. But each new name will receive a referent from the same domain of objects.

A definition of truth can now be given. The first clause of the definition describes which thoughts are expressed by atomic propositions, and the rest are rules which reflect how the thoughts expressed by complex propositions are determined by the thoughts expressed by other propositions.

Definition of truth

1. $\lceil \Pi \alpha_1 \dots \alpha_n \rceil$ is true iff $\Pi \alpha'_1 \dots \alpha'_n$.
2. $\lceil \neg \phi \rceil$ is true iff ϕ is false.
3. $\lceil (\phi \wedge \psi) \rceil$ is true iff ϕ is true, and ψ is true.
4. $\lceil (\phi \vee \psi) \rceil$ is true iff ϕ is true, or ψ is true.
5. $\lceil (\phi \rightarrow \psi) \rceil$ is true iff ϕ is false, or ψ is true.
6. $\lceil (\phi \leftrightarrow \psi) \rceil$ is true iff ϕ and ψ have the same truth-value.
7. $\lceil \forall v \phi(v) \rceil$ is true iff for every α' , $\phi(\alpha)$ is true.
8. $\lceil \exists v \phi(v) \rceil$ is true iff for some α' , $\phi(\alpha)$ is true.

We will call the schemas 1–8 *semantic axioms*. In the next section, we will see how they reflect the nature of logical truth and falsity.

It should be noted that, since the definition of truth assigns a thought to each sentence, thereby construing the sentence as a proposition, the definition of truth is also a *definition of proposition*. This replaces our initial one-line characterization of a proposition as an expression of a thought. Every sentence that is covered by a definition of truth is construed as a proposition. In other words, a proposition more appropriately characterized is what has a T-sentence in a definition of truth. Although this is ensured for all sentences in a formal language,

it is not guaranteed generally. ‘Caesar is a prime number’ and ‘colourless green ideas sleep furiously’ are sentences of English but given a sufficiently detailed theory of truth for English sentences they are unlikely to come out as propositions.

We will need some vocabulary to talk about the structures of propositions. Given our semantic axioms, the thought expressed by a complex proposition is determined by thoughts expressed by other propositions. We will call these other propositions the *bases* of the complex proposition. If a base is discernible by a single semantic axiom, like a conjunct of a conjunction, we will call it an immediate base. And if a base of a proposition does not itself have propositions as its bases, we will call it an ultimate base. All propositions have only atomic propositions as their ultimate bases.

Let us see how our definition of truth respects Convention T. Each proposition is either atomic or complex. The atomic propositions have T-sentences by axiom 1. Every other proposition receives a T-sentence from its immediate bases by axioms 2 to 8. The immediate bases are themselves either atomic or complex. And so, we run the same reasoning again. Since all complex propositions have atomic propositions as their ultimate bases, we will eventually have T-sentences for all propositions. Quantified propositions deserve special attention, however. The semantic axioms for them call for generalization over objects in the domain. For each such object, we can introduce a name into the object-language which refers to it. It does not matter which name gets introduced for the object. There will be a proposition where that name occupies the place indicated by the quantifier. We can reason in general terms about the bases of these propositions. To illustrate, consider the proposition ‘ $\forall xFx$ ’. Let the object-language relational terms be included in the meta-language. From axiom 7 we get:

‘ $\forall xFx$ ’ is true if and only if for every x , ‘ $F\alpha$ ’ is true, where α refers to x

Given axiom 1, the simple names in the object-language can be replaced with the meta-language names which are used to assign referents to them. We thus arrive at:

‘ $\forall xFx$ ’ is true if and only if for every x , Fx .

The reasoning is exactly analogous with existentially generalized propositions. Thus, we have managed to handle quantification without defining truth in terms of satisfaction. The problem concerning logical truths that we found in Tarski’s definition is thereby avoided.

In addition to avoiding the problem, the Fregean definition is also considerably simpler than the Tarski-style definition given in terms of satisfaction, although we must keep in mind

that we are working with a general understanding of names as signs that refer to objects. In fact, although the Tarski-style definition is the officially accepted definition, we rarely keep track of the distinction between truth and satisfaction in practice. This commonly leads to talk of truth relative to a variable-assignment, which violates Convention T. In T-sentences, truth is not relativized to variable-assignments or anything else. The Fregean definition gives us the desired simplicity without this violation.

There is also a straightforward translation from our new definition of truth back to Tarski's. As Dummett notes, when free variables are assigned values by a variable-assignment, each one of them is given "the effective status" of a name (*Pol*, p. 16). By assigning different values to a variable, we turn it into different names. When variables are given the status of names, all formulas are thereby given the effective status of sentences. They become expressions of thought. In other words, by fixing a variable-assignment, we turn formulas into propositions. However, when we assess the truth or falsity of a quantified proposition, we allow the variable that is to be bound by the quantifier "to run through all the individuals in the domain" (p. 17). Given our new way of thinking about variables, we are at that moment effectively considering different names for the place occupied by the variable. That is when the free variable is treated as an empty place in Frege's sense. And the formulas that contain the variable are then treated as Frege's concept-expressions. A free variable in Tarski's framework is thus playing a dual role. Under a variable-assignment, it is a name. While we are considering all variable-assignments which are its variants it is a place for a name. It should be noted, though, that although this is a way to salvage the *notation* that Tarski uses, it is not a way of saving his definition of truth in terms of satisfaction. Talk of satisfaction, if we preserved it, would just be another way of talking about truth.

8. Logical Truth

A proposition is a logical truth just in case it is true by virtue of the nature of truth. The nature of truth is captured in our semantic axioms. A proposition is thus a logical truth just in case the truth of the axioms is sufficient for the truth of the proposition. Similarly, a proposition is a logical falsity just in case the truth of the axioms is sufficient for the falsity of the proposition. For illustration, we will go through two simple examples.

First, let us prove that ' $Fa \vee \neg Fa$ ' is a logical truth. Suppose that ' Fa ' is true. Then, given our axiom for disjunctions, ' $Fa \vee \neg Fa$ ' is also true. Now suppose that ' Fa ' is not true. Then, given our axiom for negations, ' $\neg Fa$ ' is true, and thus, given our axiom for disjunctions again, ' $Fa \vee \neg Fa$ ' is still true. Thus, ' $Fa \vee \neg Fa$ ' is true regardless of the truth-values of atomic propositions. Its truth can be demonstrated by the axioms alone. It is thus a logical truth, a proposition that is true solely by virtue of what truth is.

Similar reasoning can be used to prove that ' $Fa \wedge \neg Fa$ ' is a logical falsity. Suppose that ' Fa ' is true. Then, given our axiom for negations, ' $\neg Fa$ ' is false, and thus, given our axiom for conjunctions, ' $Fa \wedge \neg Fa$ ' is false. Now suppose instead that ' Fa ' is false. Then, given our axiom for conjunctions, ' $Fa \wedge \neg Fa$ ' is false. So, ' $Fa \wedge \neg Fa$ ' is false regardless of the truth-values of atomic propositions. Its falsity can be demonstrated by the axioms alone. It is thus a logical falsity.

In the proofs, we are talking about the language and not in it. What we want, of course, is insight about logical truth in the language that we use, not in another language. But, since we have an understanding of the truth of our meta-language propositions, a definition of truth like the one we gave for the formal language can be given also for the meta-language. The definition would then be given in a new meta-language. It does not matter how a proof runs that shows of a given proposition that it is a logical truth nor which language it is given in. It does not even matter whether a proof can be given at all. The reasons why giving a proof might not be possible need not have to do with whether the truth of the proposition depends on the truth-values of atomic propositions. This can be when the object-language has further propositions characterized by further semantic axioms. Generalizations over concepts are an example. We have avoided those for the sake of simplicity. The proofs are only to illustrate that some propositions have truth-values by virtue of what truth is. Those are the propositions of logic. How or whether it can be shown of a given proposition that it is a proposition of logic is a separate matter.

One feature of propositions of logic is that their truth-values do not depend on which thoughts are expressed by the atomic propositions and thus on which referents the simple terms have. Hence the logical independence of atomic propositions. Still, for there to be truth-values at all for propositions, there need to be atomic propositions; that is, atomic sentences need to be expressions of *some* thoughts. The independence of the truth-values of propositions of logic

from any given thoughts expressed by atomic propositions lets us define what logical propositions are in a precise manner. We can do so by introducing the notion of a *model*.

A model consists of two stipulations. One is a stipulation of referents to all simple terms. This ensures that each of our atomic sentences is an expression of a thought and thus a proposition. The other is a stipulation of truth-values to the atomic propositions. It is thereby a stipulation of which thoughts expressed by atomic propositions are true and which are not, that is, of which states of affairs obtain. Given our semantic axioms, the two stipulations jointly ensure that each sentence will receive a thought in a compositional manner and is either true or false. In different models, we can have different atomic propositions since our atomic sentences can receive different thoughts. Consequently, other sentences can express different thoughts in different models as well. The logical operators that we specified with our semantic axioms will remain constant across all models.

Given how we have defined a model, when we consider different models, we are thus effectively considering different propositions. But the propositions thereby considered will be of exactly the same *form*. When switching from one model to another we only replace the atomic bases of propositions by replacing the referents of terms. The form of the whole proposition, the way its truth depends compositionally on the truth of other propositions, will stay the same since the sentence will stay the same. We have only one proposition in each model which has that exact form. We can now define logical truth and falsity.

Definition of logical truth and falsity

ϕ is a logical truth if and only if, in every model, ϕ is true.

ϕ is a logical falsity if and only if, in every model, ϕ is false.

It is customary to also define what it is for an argument to be valid. We can define validity in terms of logical truth: an argument with ϕ_1, \dots, ϕ_n as premises and ψ as a conclusion is valid just in case ' $\phi_1 \wedge \dots \wedge \phi_n \rightarrow \psi$ ' is a logical truth. We may also say that the conclusion is a logical consequence of the premises, or that the premises jointly entail the conclusion.

It should be noted that, given our definitions, whether a logical truth is true is a matter not just of the specific proposition we are considering. It also concerns what other propositions the same sentence can be turned into by replacing the referents of terms. However, the *reason* why an argument will come out valid will have nothing to do with these variations. Each of the propositions we get by replacing referents of terms will have the very same form. It is the form that matters for whether the proposition is a logical truth. The role of models in our definitions

is to give us a precise way of *ignoring* the rest. In a sense, we may say that logical truth is truth in virtue of form. But the form includes logical operations which we specified in our semantic axioms and stipulated to be constant across models. Since we haven't said anything about the nature of the logical operations, seeing logical truth in that way does not amount to much. To illustrate what is allowed by our definition to be included in the form, suppose we also had the following schemas (where α is a term):

'It is surprising that ϕ ' is true iff it is surprising that ϕ is true.

' α believes that ϕ ' is true iff α' believes that ϕ is true.

For reasons to do with Frege's puzzle that we covered briefly, they cannot really be construed as semantic axioms alongside those for the logical operations. Since substitutions of equivalents for equivalents in the instances of these schemas is not guaranteed to preserve truth, these schemas could not be used to derive T-sentences. Still, they seem to have a structural similarity to our axioms and might be taken as capturing operations of a sort with propositions as their bases. Given our current set-up, these operations would also remain constant across models. Intuitively, surprise and belief are open to empirical investigation as parts of the natural world just like houses and planets. We thus haven't yet seen why logical knowledge should differ in any interesting way from regular a posteriori knowledge. We will take up the task of demonstrating that it does in the coming chapters.

VI

Wittgenstein's Fundamental Idea

In a letter of 1912 to Russell, Wittgenstein declares that whatever the proper account of logic is, “its consequences *must* be that there are NO *logical* constants. Logic must turn out to be a *totally* different kind than any other science” (*Extracts*, p. 119). This eventually culminates in the account of logic presented in the *Tractatus*. In outline, the account is well known. “Tautologies and contradictions,” Wittgenstein tells us “lack sense,” (*TLP*, 4.461) adding: “Tautologies and contradictions are not, however, nonsensical. They are part of the symbolism” (*TLP*, 4.4611). This has come to be characterized as the view that propositions of logic are empty. It is this view, if we can defend it, that will let us establish the apriority of logical knowledge.

To know is to know *what* is true. If propositions of logic are empty, then to know a logical truth is, in a sense, to know nothing. This is a negative characterization. To make something of it, we also need a positive characterization. But this requires an understanding of Wittgenstein's view that goes beyond the mere dictum that propositions of logic are empty. Since propositions of logic must reflect the laws of truth and nothing else, any account of them must be motivated solely by appeal to the nature of truth. The task before us is thus to demonstrate Wittgenstein's view without assuming anything beyond what we introduced in the previous chapter. The understanding of logic that we will gain in this chapter will then inform our understanding of the knowledge of logic in the next and final chapter.

Our starting point will be the declaration that Wittgenstein made in the letter to Russell. It is what Wittgenstein in the *Tractatus* calls his fundamental idea, or *Grundgedanke*.⁴³

My fundamental idea is that the ‘logical constants’ are not representatives; that there can be no representatives of the *logic* of facts. (*TLP*, 4.0312)

To understand the idea, we must first understand what it negates; that is, we need to understand what it would mean for logical constants *to be* representatives and what reason there might be for supposing that they *are* representatives. We will see that the *Grundgedanke*, in essence, just is Wittgenstein's account of the propositions of logic. For its defence, we will turn to an idea that Wittgenstein inherits from Frege, known as the Context Principle:

⁴³ Henceforth, I will drop the customary italics from ‘*Grundgedanke*’ for readability.

[N]ever to ask for the meaning of a word in isolation, but only in the context of a proposition. (Frege, *FA*, p. xxii)

Dummett's and Anscombe's work will help us clarify what to make of it. We will see that any understanding of truth presupposes the Context Principle, and then, that the Grundgedanke is a consequence of it.

1. Denying the *Grundgedanke*

Wittgenstein's Grundgedanke, in a nutshell, is the idea that signs like ' \neg ', ' \wedge ', ' $\forall x$ ', etc. do not refer to objects, that "[t]here are no 'logical objects'" (*TLP*, 4.441).⁴⁴ Thus stated, it is initially hard to imagine anyone denying it. And indeed, Wittgenstein intends the idea to be intuitive. When it first appears in *Notes on Logic*, Wittgenstein says:

That "or" and "not" etc. are not relations in the same sense as "right" and "left" etc. is obvious to the plain man. (*NL*, MS III, p. 101)

Intuitiveness aside, the contemporary *explanation* for why the Grundgedanke is true is likely to be based on some confusions that we need to clear up.

To say that the logical constants do not refer is to say that they do not function as terms. To this it might be added that in a definition of truth the logical constants are given a very different treatment from terms. We do not stipulate referents to the logical constants but instead specify their semantic values by recursive clauses. But this is a misunderstanding of what the recursive clauses do, or equivalently, what the logical constants are. The recursive clauses specify how thoughts expressed by propositions of a given form are determined by thoughts expressed by their immediate bases. The clauses thus explicitly treat only of forms of propositions, not of the signs ' \neg ', ' \wedge ', etc.

Once we recognize that our definition of truth does not give an explicit treatment of the signs themselves, a question is likely to arise concerning the semantic values of the logical constants. The worry might be motivated by constraints of compositionality that we noted in the previous chapter, that to define truth for a language with infinitely many propositions, each

⁴⁴ It should be noted that Wittgenstein in fact has a much broader notion of logical constants in mind than we are assuming here. Wittgenstein's logical constants include, for example, propositions, forms, and expressions of truth-value. Concerning the latter, we have already seen that they are not terms since truth cannot be construed as an addition to a thought and so cannot be an object.

simple sign should have its own characteristic contribution to the thought expressed by any proposition in which the sign occurs. So, in a T-sentence like

$'p \wedge q'$ is true iff p and q

there should be for each component of the proposition a corresponding component of a thought. The sign ' \wedge ' is then naturally seen as contributing to the thought the component expressed in the meta-language by 'and'.

It is not obvious that compositionality motivates an assignment of referents to the logical constants. The definition of truth we gave earlier does handle infinitely many propositions and does so while remaining silent about the semantic values of logical constants. All that it explicitly recognizes are forms of propositions. Still, it may be thought that in giving a semantic treatment to propositions based on their form we are implicitly also assigning values to the logical constants, that in specifying the axiom for conjunctions

$'\phi \wedge \psi'$ is true iff ϕ is true and ψ is true

we are invoking thoughts with a specific component, a component we express by 'and' in the meta-language. If that is the case, then even if do not define a referent to ' \wedge ' explicitly, we still end up with an account where the sign has a referent. Every other component of the proposition has been allowed to go variable, except for ' \wedge '. And likewise, every other component of the thought is allowed to go variable except for the logical object. By assigning thoughts to propositions in this way, we end up matching ' \wedge ' with the logical object.

Further support for the denial of the Grundgedanke may be drawn from intuitive considerations concerning logically equivalent propositions. Cresswell, for example, remarks that "we may have a person so logically blind that he may believe p without believing $\sim\sim p$," adding in parentheses that "[i]f this sounds far fetched we need only introduce other functors and produce more complicated logical equivalences" (1975, p. 26). In other words, there seems to be a cognitive gap between logically equivalent propositions; that it is possible to be in doubt about the truth of one while assuming the truth of the other settled. If so, then logically equivalent propositions must express different thoughts. The right-hand sides of their T-sentences can certainly take different shapes. At least on the face of it, the only way to account for the difference in thought between, say, ' $\sim p \wedge q$ ' and ' $\sim(p \vee \sim q)$ ', is to appeal to the contributions of the logical constants occurring in the propositions. Given the Grundgedanke, this way of accounting for the cognitive gap is excluded. The only components that the thoughts

have would be those which are contributed by the atomic bases of the propositions. And ' $\neg p \wedge q$ ' and ' $\neg(p \vee \neg q)$ ' have the same atomic bases.

This implication of the Grundgedanke can lead to another confusion. The idea that logically equivalent propositions express the same thought invites a parallel with coarse-grained truth-conditions. Although the Grundgedanke may have some implications in common with a coarse-grained account of truth-conditions, it is not itself such an account. In a coarse-grained account, like the one defended by Stalnaker (1984), truth-conditions are construed in terms of *when* a proposition *is* true. Such an account implies that necessarily equivalent propositions have the same truth-condition. It is an account that aims to get the extension of the truth-predicate right and supposes it to be the end of the matter. We, on the other hand, are construing truth-conditions (or thoughts) in terms of *what it is* for a proposition *to be* true; of what the truth of a proposition consists in. Truth-conditions on this construal still have structure, with components of propositions contributing components to thoughts. What the Grundgedanke seems to imply, in other words, is that, given the *finest* truth-conditions, logically equivalent propositions have the same truth-condition.

We have now seen some motivation for rejecting the Grundgedanke, but we have yet to cover what exactly the Grundgedanke is a denial of. It states that the logical constants are not terms that refer to logical objects. But what would the logical objects be? We have already seen why they could not be truth-functions. Truth-functions are characteristic of forms of propositions. Given our treatment of propositions of the form ' $\phi \wedge \psi$ ', they are true just in case their immediate bases are true, and false otherwise. This can be characterized by a two-argument function from truth-values to truth-values. The values of these functions are the truth-values of propositions of the form ' $\phi \wedge \psi$ ', or equivalently, of thoughts expressed by those propositions. So, we must assume that conjunctions already express thoughts to speak of a truth-function that characterizes conjunctions. Since logical objects are supposed to be constituents of these thoughts, their presence is presupposed by talk of truth-functions. In other words, truth-functions cannot be logical objects because they must be the *upshots* of the role played by logical objects in thoughts.

A parallel with relational terms and their referents helps to clarify what the putative logical objects would be. Let us relax the constraints of formal syntax for readability and allow English relational terms to occur in our object-language. Given our axiom for atomic propositions, we have the T-sentence

'a loves b' is true iff *a loves b*

Here, the relational term 'loves' contributes the love relation to the thought, which the proposition says holds between *a* and *b*, in that order. The proposition is thus true just in case *a loves b*. Similarly, it might be supposed that in a T-sentence

'p ∧ q' is true iff *p* and *q*

the logical term '∧' contributes a relation of being jointly true, which the proposition says holds between two thoughts, the thought that *p* and the thought that *q*. The proposition would be true just in case the relation holds between the thoughts, that is, when both thoughts are jointly true. Thus, a truth-function would be determined. Likewise, just as 'red' in '*c is red*' contributes the property of redness which is characteristic of all and only the red things, '¬' in '¬*p*' might be said to contribute the property of not being true which holds of all and only the thoughts which are not true. Like with the relation contributed by '∧', this property would give us the expected truth-function which maps truth to falsity and falsity to truth. The logical objects, in other words, would be properties of thoughts and relations between thoughts, and the signs for logical operations would be names of these properties and relations. And so, in a quite straightforward sense, 'or' and 'not' *would* be "relations in the same sense as "right" and "left"" (NL, p. 101).

The existence of logical objects has implications for the nature of logical truths. First, logical truths would express thoughts. They would concern logical properties and relations in roughly the same way as '*a loves b*' concerns love. Although '*Fa ∨ ¬Fa*' might still be said to involve reference to *F* and *a* which would thus be components of the thought expressed, the truth of the proposition would depend, not on what is characteristic of *F* and *a*, but on what is characteristic of the thought that *Fa*. The thought that *Fa* would be acting like a term in a relation contributed by the expression '*... ∨ ¬ ...*'. And whether the result is true would depend on whether the thought that *Fa* stands in that relation to itself. As Ramsey (1927) characterizes this view, all logically complex propositions, whether logically true or not, would be "similar to atomic ones except that the relations they assert are logical instead of material" (p. 146). And so, on this view, as Ramsey points out, "every proposition is ultimately affirmative, asserting a simple relation between simple terms, or a simple quality of a simple term" (ibid.). But this invites a question. Why is a proposition like '*Fa ∨ ¬Fa*' *logically* true? If the contingent proposition '*Fa ∨ ¬Gb*' is true, then the same relation would hold between the thought that *Fa* and the thought that *Gb*. In this way, there is a straightforward parallel with '*a loves a*' and '*a loves b*'. This is the worry that Ramsey expresses when he says: "I find it very

unsatisfactory to be left with no explanation of formal logic except that it is a collection of ‘necessary facts’” (ibid.). The logical difference between the truth of ‘ $Fa \vee \neg Fa$ ’ and the truth of ‘ $Fa \vee \neg Gb$ ’ could consist only in that the truth of the former is necessary, guaranteed in some way by the nature of the relation referred to by ‘ $\dots \vee \neg \dots$ ’. But the proposition would no more tell us about this nature than ‘ a loves a ’ tells us about the nature of love.

It is not difficult to see how this construal of logical truths can invite a way of knowing approach to logical knowledge and how this in turn reasonably leads to scepticism about the apriority of such knowledge. Given the existence of logical objects, logic as a discipline would be the study of the nature of relations of a particular sort. There is no reason for thinking logic to be fundamentally different from any other science. We can conduct a similar investigation, for example, into social relations between people. There will be some constraints on methodology due to difference in subject-matter just as there is in a study of distant heavenly bodies or the subconscious. But such constraints in logic as in all other fields are something for each individual investigator to explore and, if possible, overcome. To maintain that logical knowledge is a priori, given this view, would be to impose methodological demands of the sort that Feyerabend has warned us are liable to lead to bad science and unhappy scientists. The primary reason for being sceptical about claims of apriority in logic would thus not be any positive reason for thinking that logical knowledge is empirical, but simply that it is nobody’s business to dictate how logical objects ought to be investigated.

2. Truth Intrinsic to the Proposition

We have now seen what are the broader implications of denying the Grundgedanke and have *some* idea of what Wittgenstein had in mind when, after telling Russell that there cannot be logical constants, immediately follows it with: “Logic must turn out to be a *totally* different kind than any other science” (*Extracts*, p. 119). But we can elaborate it further.

The truth of a logical truth does not depend on the truth-values of its atomic bases. If there are no logical objects, that is, if the Grundgedanke is true, then there is nothing at all for their truth to depend on. In contrast to the truth-values of all other propositions, the truth-values of propositions of logic would then not be *answerable* to how things are. But, given our definition of truth, for a proposition to express a thought is for it to be true just in case the thought is. And a thought just is a way for things to be, something that, if true, is a fact. So,

given the Grundgedanke, propositions of logic would not express thoughts. There is nothing that the world must be like for a logical truth to be true. This explains why Wittgenstein characterizes propositions of logic as empty and why logical knowledge would be knowledge about nothing. But we have yet to see how to spell all of this out in positive terms.

What does it *mean*, it may reasonably be wondered, for something to be true or false when this is not a reflection of what is or is not the case? Logical positivists, who adopted Wittgenstein's view of logic, at least in outline, addressed just this question. Crediting Wittgenstein, Carnap maintained that logical truths are true and logical falsities false "by virtue of their form" and that they "say nothing about reality" (1932, p. 76). Echoing Wittgenstein (*TLP*, 4.461), Ayer likewise notes that a proposition like 'either some ants are parasitic or none are' "provides no information whatsoever about the behaviour of ants, or, indeed, about any matter of fact" (1952, p. 79). Given how we construe thoughts, to say that a proposition says "nothing about reality" or that it provides no information about "any matter of fact" just is to say that it does not express a thought. But to this negative characterization, Carnap and Ayer also add a *positive* characterization.

Carnap suggests that logical truths "are not themselves factual statements, but serve for the transformation of such statements" (1932, p. 76). Roughly, Carnap's idea is that logical truths encode proof-rules. For example, the purpose of a sentence of the form

$$\neg(p \wedge q) \leftrightarrow \neg p \vee \neg q$$

according to this idea, is to record the fact that we can freely switch between ' $\neg(p \wedge q)$ ' and ' $\neg p \vee \neg q$ '. Ayer gives a similar positive characterization, proposing that logical truths "enlighten us by illustrating the way in which we use certain symbols" (1952, p.79). For example, by uttering a sentence like 'if all Bretons are Frenchmen, and all Frenchmen Europeans, then all Bretons are Europeans', Ayer says,

I am showing that in the statement that all Bretons are Frenchmen, and all Frenchmen Europeans, the further statement that all Bretons are Europeans is implicitly contained. And I am thereby indicating the convention which governs our usage of the words "if" and "all." (ibid. 80)

The purpose a proposal like Carnap's or Ayer's is not immediately clear. On one way of understanding it, the proposal simply aims to explain why we sometimes utter logical truths despite them expressing no thoughts. But any such explanation will be an empirical hypothesis concerning the pragmatics of uttering these propositions in communication. And taken in this

way, Carnap's and Ayer's proposal is surely too narrow. When Pilate said, "What I have written, I have written", he could hardly be taken as educating the audience about the use of the word 'what' (or "O" in Greek). But more importantly, in logic, there is no *need* for an account of the pragmatics of uttering logical truths.

There is, however, another way of taking the positive characterization. We might take it as a claim that the pragmatic story *exhausts* the role of propositions of logic in the language—that there is no semantic account of them to give, or that the pragmatic account just is the semantic account. It is unlikely that Carnap or Ayer really maintained this, but it would be a mistake if they did. The proposal would then be that propositions of logic are not in the business of stating truths and falsities in the normal sense of 'true' and 'false'. The word 'true', when applied to a logical truth, will not mean *true* in the sense in which 'snow is white' is true just in case snow is white. But a logical truth *must* be true in just this sense. The semantic contribution of each expression involved in a proposition of logic is the same as in all other propositions. It is because of this that propositions of logic, like all propositions, receive T-sentences in a general definition of truth. This is the sole reason for recognizing them as special. We set out to define truth in the normal sense, with Convention T as a criterion of adequacy, and found that in *that* sense of 'true' some propositions are true independently of the truth-values of their atomic bases. We might go on to posit that a proposition like ' $p \vee \neg p$ ' serves, for example, as a linguistic lesson and indicate which lesson it is by assigning the proposition some value, say, *P* (for 'permission'). But that value can only be an *addition* to the *truth*-value that the proposition has. It is simply not up to us to withhold truth from logical truths or falsity from logical falsities.⁴⁵

So, whatever we make of the Grundgedanke, propositions of logic must have genuine truth-values. But contrary to intuition this does not mean that they must express thoughts, that is, have truth-conditions. Logical truths and falsities certainly have T-sentences. But we need not suppose that all T-sentences convey which thoughts are expressed by propositions. This does not require any significant revision in our understanding of T-sentences. A T-sentence is itself a proposition in the metalanguage. Its right-hand side conveys a thought, when it does,

⁴⁵ For a different objection to accounts like Carnap's and Ayer's, see Quine 1936, 1960. Quine's objection in a nutshell is that, since there are infinitely many propositions of logic, any *assignment* of dummy values to them must assume the truth of some logical truths to have the required generality. Quine's objection is targeted more specifically to the idea that logical truths are true by convention. It is because convention is established in a finite time that the question arises concerning how infinitely many logical truths could be covered by it. The objection I have given above does not require the involvement of convention or anything else of the sort.

by *expressing* that thought in just the way that the object-language proposition does. If we deny that propositions of logic express thoughts, we deny it also for the propositions of logic formulated in the metalanguage and thus for the right-hand sides of the T-sentences of propositions of logic. There is thus no immediate problem from formal semantics to the idea of truth-values without thoughts or truth-conditions.

We began this section noting that, if the Grundgedanke was true, then the truth-values of propositions of logic would not be answerable to how things are. We then asked, in effect: what *are* they answerable to? We can now see why this question was confused. There is no theoretical gap left by denying that logical truths express thoughts. The definition of truth by itself gives the only positive characterization of propositions of logic that we could ask for. Given a compositional definition of truth, logical truths and falsities are bound to have truth-values. But a definition of truth is also a definition of proposition. Thus, *if* there are no logical objects, then propositions of logic have truth-values solely by virtue of the sort of propositions that they are. They would, in other words, be *intrinsically* true or false. But still, they would be true or false in just the sense that ‘snow is white’ is true if and only if snow is white. It is in the face of this result that one might push back on the idea that T-sentences need not convey which thoughts are expressed by propositions.

The assumption that thoughts map one-to-one to right-hand sides of T-sentences could be seen as an attempt to capture a deeper idea concerning the nature of truth. As David Armstrong puts the idea, “a truth, any truth, should depend for its truth for something ‘outside’ it, in virtue of which it is true” (2004, p. 7). Armstrong admits he has no argument to give for this but adds: “My hope is that philosophers of realist inclinations will be immediately attracted to the idea” (ibid.). If every truth needs to have “something ‘outside’ it” to make it true, then the right-hand sides of T-sentences *must* be read as expressing thoughts. As stated, the objection is simply an outright rejection of the Grundgedanke. So, a proper response to it would be a defence of the Grundgedanke, which we will take up below. Still, we can explore how well-founded the objection is. First, we can note that it concerns the truth of propositions, not the truth of thoughts. A thought is not true by virtue of “something ‘outside’ it” but is itself the “something” outside. We are thus in the realm of a definition of truth for sentences. Tarski (*SCT*, p. 343) rejects all one-liners such as

“The truth of a sentence consists in its agreement with (or correspondence to) reality.”

“A sentence is true if it designates an existing state of affairs.”

not on the grounds of any metaphysical views but because “none of them is sufficiently precise and clear,” noting that “a more precise expression of our intuitions” is required (ibid.). This then leads to a formulation of Convention T, the requirement that a definition of truth for a language should yield a T-sentence for every sentence of the language. The idea that truth requires “something ‘outside’ it” is to be rejected on the very same grounds, as only a rough attempt to capture the gist of Convention T. It is ultimately Convention T itself that matters not a one-line summary of it. If a theory that adheres to Convention T leaves room for the truth of a proposition without “something ‘outside’ it,” then this just means that the one-line summary was not as precise and clear as we initially supposed.⁴⁶

We have now covered the Grundgedanke and its implications for logic. If the Grundgedanke is false, then propositions of logic describe logical relations between thoughts and depend for their truth or falsity on whether those relations hold. If the Grundgedanke is true, then truth-values are intrinsic to propositions of logic and not answerable to anything. The rest of this chapter is an extended defence of the Grundgedanke.

3. The Context Principle

In *Foundations of Arithmetic* (p. xxii), Frege famously adopts three principles to guide his investigation. We covertly adopted the first principle, “always to separate sharply the psychological from the logical, the subjective from the objective,” in chapter III when we rejected the way of knowing conception of the a priori-a posteriori distinction. Taking a lesson from Ramsey’s work, we adopted the third principle, “never to lose sight of the distinction between concept and object,” noting how this principle goes deeper than Frege plausibly realized himself. We have yet to cover the second, known as the Context Principle:

[N]ever to ask for the meaning of a word in isolation, but only in the context of a proposition.

⁴⁶ A further worry one might have is that, in claiming that propositions of logic do not express thoughts, we are misconstruing Frege’s notion. Frege introduces the notion of a thought by defining a thought as that to which the question ‘is it true?’ is applicable. But if the question is applicable, it can be answered in either of *two* ways. It can receive an affirmative or a negative answer. Only one of the answers is correct, but either one can be given. So, Frege’s notion of a thought, *pace* Frege, involves more than the possession of a truth-value. It involves the possibility of having *either* truth-value. For propositions of logic, given the Grundgedanke, there is only one. And so, strictly speaking, they do not count as expressing thoughts in Frege’s sense.

Wittgenstein adopts the principle in the *Tractatus*.

Only propositions have sense; only in the nexus of a proposition does a name have meaning. (*TLP*, 3.3)

Dummett elaborates the principle as a statement that the primary unit of meaning in language is the proposition; that, if “we know everything that needs to be known about a word in order to determine the condition for the truth of any sentence [proposition] in which it occurs, we know the reference of the word” (*PoL*, p. 195). The idea in other words is that meaning is to be understood solely in terms of contributions to the determination of thoughts. And the reason, Dummett maintains, is that only if we understand the meanings of words *as* contributions to thoughts from the start could we explain how the words could jointly determine thoughts. Otherwise, we are back to speaking “vaguely of words as conveying ideas” in the way familiar from the traditional empiricism of Hume and Locke (*ibid.*). Understood in this way, the Context Principle is the now widely accepted idea that forms the foundation of formal semantics; that investigation of meaning must go via an investigation of thoughts (truth-conditions). All of this is surely right, but it does not go far enough.

A different explanation is given by Anscombe in her introduction to Wittgenstein’s *Tractatus* (*IWT*, Ch. 4). Since Anscombe is discussing Wittgenstein’s picture theory, she gives her explanation in terms of pictures. A picture, as Wittgenstein uses the term, is something that represents a situation (*TLP*, 2.11), or, in our preferred terminology, something that expresses a thought. Propositions are then introduced as “logical pictures,” pictures where there is only the bare minimum required for the representation of a situation (2.18–81). Since “[e]very picture is *at the same time* a logical one” (2.182), we may read Anscombe’s explanation as concerning propositions directly.

Anscombe begins by noting a peculiar intuitive feature of reference, namely that the idea of reference seems to make no sense unless it is in connection to representation.

Suppose I said: ‘That door stands for Dante and that table for Bertrand Russell’. My audience would, if anything, look at me enquiringly and say: ‘Well?’ And here ‘Well?’ means ‘Do something to shew the point of this’; and *that* means ‘Let something else come into such a relationship with this door, or again with this table, that the terms in relation represent something.’ (*IWT*, pp. 66–67)

The point concerning doors and tables carries over to sounds and marks on paper. What Anscombe takes it to illustrate, as is evident from her construal of the audience reaction, is that,

unless we combine our declarations of reference with something further so that “the terms in relation represent something,” we do not really refer. Although, with respect to time, it may first be said ‘let this sign stand for this thing and that for that thing,’ its success is conditional on whether the signs will be employed as part of something that could convey a thought. “[U]nless it is that,” Anscombe says, “it is not anything at all” (p. 66). This leads to Anscombe’s formulation of the Context Principle:

Only in the connections that make up the picture can the elements of the picture stand for objects. (p. 67).

As she elaborates it, a picture must have two features: “first, the relation between the elements of the picture; and second, the correlations of the elements in the picture with things outside the picture,” and, Anscombe adds, “the first feature must belong to a picture before the second one can” (p. 68). The intuitive examples are of course only for illustration. But before we turn to Anscombe’s justification of the Context Principle, we should note in what respect Anscombe’s reading of the principle differs from Dummett’s.

The main difference is not as great as might initially appear. It is mainly one of emphasis. If we put it in terms of reference, as Anscombe does, then Dummett’s focus is on what it is for a term to have the *specific* referent that it has, whereas Anscombe reads the principle as concerning what is required for a term to be capable of *any* reference at all. The question ‘*what* does a term refer to?’ is not where Anscombe’s focus is. However, to make sense of Dummett’s reading, we *do* need to assume that terms have referents only as components of propositions. Only then could their referents be fully determined by truth-conditions. So, the two readings of the Context Principle are at least very closely related.⁴⁷ But it is Anscombe who points to the reason why it holds.

The main reason that Anscombe gives is the point we considered earlier, namely that “[y]ou cannot divide ‘The cat is often drunk’ into ‘The cat is of’ and ‘ten drunk’, although each part could be significant” (*IWT*, p. 75). We noted earlier that the sense in which we cannot divide the proposition that way is that a grasp of the thought expressed by the proposition is part of a general understanding of truth. This general understanding, since it is captured by Convention T, presupposes seeing propositions as having compositional structure where each

⁴⁷ Dummett seems to be unaware of the parallel. This is evident from his own gloss on Anscombe’s cases, that “[u]ttering an expression *which refers* to an object has in itself no significance” (*PoL*, p. 297, emphasis added), and that without some implied purpose, “assigning a bearer to a name would be merely an empty ceremony” (*PoL*, p. 495). Anscombe does not say it is an empty ceremony. She says that “it is not anything at all” (*IWT*, p. 66).

component of a proposition makes its distinctive contribution to the determination of a thought. And, as it happens to be in English, the thought expressed by ‘the cat is often drunk’ is not determined by the values of ‘the cat is of’ and ‘ten drunk’. But how does this yield the Context Principle? Anscombe continues:

as Wittgenstein would say, the first expression’s standing to the left of the second is not what signifies in our sentence, and (3.314) ‘An expression has reference only in the context of a proposition’. (It was of course on this pronouncement, and on that of Frege, repeated by Wittgenstein . . . that I modelled my statement about pictures: ‘It is only in the connections that make up the picture that its elements stand for objects.’) (ibid.)⁴⁸

So, by “the connections that make up the picture” Anscombe has in mind the sort of connections that there are between objects in a situation, as in: “the first expression’s standing to the left of the second.” We saw in the previous chapter why this must be so, why ‘*Fa*’, for example, is to be understood as ‘*F*’s standing to the left of ‘*a*’ rather than as the result of putting ‘*F*’ before ‘*a*’. Unless we construe our sentences as situations, our definition of truth for sentences could only capture truths which have been formulated. So, what Anscombe says is that only when signs themselves form a situation can they stand for something else.

An example helps to illustrate Anscombe’s point. Suppose we had ‘*GF*’ and ‘*F*’ among our relational signs. In that case, our object-language would have ‘*GFa*’ and ‘*Fa*’ as propositions. Why does the ‘*F*’ in ‘*GFa*’ not refer to an object, even though in ‘*Fa*’ it does? The answer is seen when we unpack the propositions:

‘*GFa*’ is to be read as ‘*GF*’ followed by ‘*a*’

‘*Fa*’ is to be read as ‘*F*’ followed by ‘*a*’

In a literal sense, ‘*GFa*’ does not have ‘*F*’ as a *component*. In it, ‘*F*’ does not stand to other elements in any way. If ‘*F*’ was a component of ‘*GFa*’, we would have no way to exclude it from contributing an object to the thought expressed since, for a definition of truth to cover an infinity of sentences, an expression must make the same contribution in *every* proposition in which it occurs. So, in cases of merged notation, such as in our ‘*GFa*’, it must first be decided which sentence-situation we take the signs to occur in before we can settle whether they stand

⁴⁸ When citing *TLP* 3.314, Anscombe is relying on her own translation. The passage in original German reads: “Der Ausdruck hat nur im Satz Bedeutung.” In both Ogden’s translation and the translation by Pears and McGuinness, the word ‘Bedeutung’ is translated as ‘meaning’, thus making it harder to see Anscombe’s point in the passage. Of course, it does not ultimately matter since “[a] name means an object” (*TLP*, 3.203).

for objects and thus whether they are terms. The proposition cannot be construed as “a blend of words,” as Wittgenstein puts it; “[a] proposition is articulate” (*TLP*, 3.141).

There is an objection that examples like this as well as Anscombe’s own example might invite. All that they show, it might be argued, is that there is a need for notational hygiene, a need to avoid cases like our ‘*GFa*’ and ‘*Fa*’. There is, the objection continues, nothing to do about it in natural languages since it is an empirical matter what notation is used in them. As theorists, we cannot prescribe notation to speakers. Because of this, the Context Principle needs to be assumed in natural language semantics. But it must be assumed only as a *methodological* principle that helps us deal with a specific kind of notational ambiguity. The notation of our formal language, on the other hand, is up to us. And so, the suggestion goes, we can ensure that we do not need to assume the Context Principle in a formal language.⁴⁹

What the objection fails to take into account is generality. For generalizations over arbitrarily large domains to be possible, there can be no fixed list of potential terms for objects and thus no fixed list of shapes for the notation of terms.⁵⁰ So, a definition of truth must cover all instances of a generalization *independently* of our choice of notation for terms. Of course, we can make sure to always write our terms so as to avoid ambiguity whenever we need a term, but this is beside the point. What the need to account for generalities shows is that we cannot construe *the shapes themselves* as terms. For a definition of truth to account for generality, a term must be specifiable by its role alone and independently of any notation.

What then is a term if not a shape? Dummett’s reading of the Context Principle can help to answer this. As Dummett emphasized, unless we already construe the referent of a term *as* a contribution to thoughts expressed by propositions in which the term occurs, we will not be able to explain how a term having a referent concerns thoughts. The problem is not that we couldn’t explain how complex expressions have complex meanings. On the contrary, that is all that we *could* say. As Dummett characterizes the time before Frege:

⁴⁹ I am making the methodological reading of the principle sound less significant than it is. Consider a formal language that makes use of shapes ‘*I*’ and ‘*F*’. In the Cambria Math font that they are written in, the latter has the former as a component. Without the Context Principle, we would have no way of blocking the occurrence of ‘*I*’ from contributing to the thought expressed by ‘*Fa*’. It is of course silly to suppose that ‘*I*’ *might* be playing such a role in ‘*Fa*’. But that just goes to show the extent to which the Context Principle as a methodological principle is required.

⁵⁰ Tarski’s notation of free variables hides the need for an open-ended range of names by giving us a notation that can remain constant across instances of a generalization—we can go from ‘ $\forall xFx$ ’ to ‘ Fx ’ under one or another variable-assignment—but this does not mean that the notation is fixed beforehand. It is still up to us to write the variable the same way for every instance. A variable under an assignment has, as Dummett says, “the effective status” of a name (*PoL*, p. 16) and thus the same constraints apply.

Earlier writers had spoken vaguely of words as conveying ideas, and complex expressions as conveying complex ideas compounded out of the ideas conveyed by the constituent words: a sentence then becomes a particular form of complex expression, and the unique and central character of this category of linguistic expression is left entirely unaccounted for . . . (*PoL*, p. 195)

In other words, we could explain complex meanings but not why those complex meanings should be playing the unique role of thoughts. A thought is not a mere complex of components. The parts of a thought must hang together to make the question ‘is it true?’ applicable. It is only if we “come by the parts of a thought by analysing the thought” (Frege, *PW*, p. 253), that is, by construing objects *as* parts of the thought, that we have an account of that. In Tractarian terms, “[i]t is essential to things that they should be possible constituents of states of affairs” (*TLP*, 2.011).

In an exactly parallel fashion, we could not explain how simple terms come together to form a sentence—something that *can* express a thought—unless we come by the terms by analysing the sentence. All we would get are combinations of words. Just as an object referred to by a term is to be found only in a thought, the term itself is to be found only in a sentence. And as we have seen, a sentence itself must be construed as a situation. So, strictly speaking, we do not even need an account of terms that *parallels* that of objects. Terms *are* objects. Dummett’s point thus generalizes to terms. A sentence-situation, when formulated, does involve shapes of a particular sort: ‘*Fa*’ is to be read as the *sign* ‘*F*’ followed by the *sign* ‘*a*’. But once we individuate terms by their role in the sentence-situation, the specific characteristics of the shapes are irrelevant to the constitution of the situation. All that we need to consider is what is required for the sentence-situation to represent.

A proposition possesses essential and accidental features.

Accidental features are those that result from the particular way in which the propositional sign is produced. Essential features are those without which the proposition could not express its sense. (*TLP*, 3.34)

This is enough to ensure that our language has all the terms it needs to account for generality regardless of domain. Terms are simply not individuated by their typography, which would require us to identify them separately from our definition of sentence. They will still have shape, colour, sound, etc. but only as accidental features. The same term could be written

differently. And we can remain neutral about *how* it is written.⁵¹ We individuate terms instead by their role in the determination of thoughts expressed by propositions which themselves are defined schematically. But this means that terms exist only *as components* of propositions. As Wittgenstein puts it:

It is impossible for words to occur in two different ways, alone and in the proposition.
(*TLP*, 2.0122, Ogd.)

In order to recognize the symbol in the sign we must consider the significant use. (3.326,
Ogd.)

We thus cannot treat the shape ‘Elizabeth Anscombe’ as a name in the sense of something that could refer unless we specify some sentence *in which* it could refer. That is, we would need to specify a certain significant use for it.

This explains why the Context Principle, as Anscombe reads it, holds. It is only in a sentence-situation that we can identify terms—things of the sort that could stand for something. And only then could we posit referents to them. When we say, “That door stands for Dante and that table for Bertrand Russell,” we are not employing the door and the table as components of a situation in the way that is required for them to count as terms. It is only when we consider the door and the table as part of a single situation—something that could serve as a sentence—that we can recognize them *as* terms in the whole and assign referents to terms. In other words, we must “[l]et something else come into such a relationship with this door, or again with this table, that the terms in relation represent something” (*IWT*, pp. 66–67).

We noted earlier that the difference between Dummett’s and Anscombe’s reading of the Context Principle is primarily in emphasis. We can now see that the difference is *solely* in emphasis. Dummett’s focus is on the constitution of a thought whereas Anscombe’s is on the constitution of a sentence. But both are building on the same idea, that of “giving pride of place to the content of the word ‘true’” (Frege, *PW*, p. 253). There is thus only one Context Principle but with many implications.

⁵¹ The idea of expressions being individuated in terms of their semantic features but also having accidental features helps to alleviate a tension that contemporary semanticists sometimes feel. For example, Seth Yalcin recognizes that “[i]t is natural to theorize with a use of ‘expression’ according to which expressions have semantic values necessarily—indeed, have the semantic values that they in fact have necessarily,” but immediately goes on to qualify this, noting that “expressions should be individuated *in part* by their semantic values” (2015, p. 36, emphasis added). One is tempted to ask: what the other part is? It cannot be the sounds or shapes since those vary from user to user. And there seem to be no other candidates. By recognizing accidental features of expressions, Wittgenstein provides a clear solution to this conundrum.

4. In Defence of the *Grundgedanke*

With the Context Principle in place, we can prove the *Grundgedanke* from it. To quote the passage in full:

The possibility of propositions is based on the principle that objects have signs as their representatives.

My fundamental idea is that the ‘logical constants’ are not representatives; that there can be no representatives of the *logic* of facts. (*TLP*, 4.0312)

We have seen what significance the preceding statement has. For signs to be representatives of objects, that is, for signs to refer, they must be considered in a sentence-situation, an articulated network of signs. Our task is to explain how this implies the *Grundgedanke* that comes immediately after.

To avoid assuming any treatment of the signs for logical operations, the proof needs to be formulated as concerning logical operations on atomic propositions. As we construed atomic propositions in the previous chapter (and as Wittgenstein himself construes them), an atomic proposition is a concatenation of terms with each term standing for an object. The whole expresses a thought that can be parsed into objects. To repeat, our axiom for atomic propositions is the following:

$\lceil \Pi \alpha_1 \dots \alpha_n \rceil$ is true iff $\Pi' \alpha'_1 \dots \alpha'_n$.

That the *Grundgedanke* and Wittgenstein’s account of logic in general is implied by an account of atomic propositions is suggested by a famous passage in a letter of 1912 from Wittgenstein to Russell, where Wittgenstein reports on the progress of his work concerning logical constants.

I believe that our problems can be traced down to the *atomic* propositions. This you will see if you try to explain precisely in what way the Copula in such a proposition has meaning. I cannot explain it and I think that as soon as an exact answer to this question is given the problem of “v” and of the apparent variable will be brought *very* near to their solution if not solved. I now think about “Socrates is human” (Good old Socrates!). (*Extracts*, p. 121)

To run the proof from our account of atomic propositions to the *Grundgedanke*, we also need to pick a logical operation for illustration. We will pick negation since it is Wittgenstein’s most used example; but our conclusion will hold generally for all signs for logical operations, as we

will see. So, we are immediately concerned with propositions of the form $\lceil \neg \Pi \alpha_1 \dots \alpha_n \rceil$. Since the specific terms do not matter, we may pick ‘*Fa*’ as our atomic proposition for brevity.

To the set up the proof, there is one last point about logically complex propositions that we need to note. In the *Notebooks*, Wittgenstein writes:⁵²

If a picture presents what-is-not-the-case . . . , this only happens through its presenting *that* which *is* not the case.

For the picture says, as it were: “*This* is how it is *not*”, and to the question “*How* is it not?” just the positive proposition is the answer. (*NB*, p. 25)

In other words, a negation ‘ $\neg p$ ’ must involve an expression of the thought that it negates. Otherwise, ‘ $\neg p$ ’ will not *be* a negation. What is negated by ‘ $\neg p$ ’ is just what is expressed by the proposition ‘*p*’. So, for ‘ $\neg p$ ’ to be the negation of ‘*p*’, ‘*p*’ in ‘ $\neg p$ ’ must express a thought. Therefore, ‘*p*’, as it occurs in ‘ $\neg p$ ’, must have whatever characteristics that are required for it to be a proposition. As Wittgenstein puts it:

Only a finished proposition can be negated. (*NB*, p. 25)

Every proposition must *already* have a sense: it cannot be given a sense by affirmation.

Indeed its sense is just what is affirmed. And the same applies to negation, etc. (*TLP*, 4.064)

The same point is emphasized by Anscombe; that in a negation, “one must convey *what* situation one is saying does not exist, and this will be conveyed precisely by the picture depicting that situation” (*IWT*, p. 69); that is, the proposition negated must express a thought if the negation is to be a negation *of* that thought.

This is a familiar feature of negation. We must first recognize what the negated proposition says before we can recognize what the negation says. We understand what ‘it is not raining says’ by grasping the thought expressed by ‘it is raining’ and recognizing the whole proposition as negating that thought. Our overview of the Grundgedanke from before itself took this same form. The Grundgedanke is a negation. To explain what it says, we first had to explain what it negates.

It is of course not merely an intuitive feature of negation. Our definition of truth explains why this must be so. A thought expressed by a negation is determined at least partly by a

⁵² The same idea occurs already in *Notes on Logic*: “however . . . “not-p” may be explained, the question what is negated must have a meaning” (*NL*, p. 94). In *Notes on Logic*, Wittgenstein assumes a different conception of meaning from the one developed in the *Notebooks* and presented in the *Tractatus*. So, the statement is not strictly speaking the same. But it is clear that Wittgenstein is aiming for the same basic idea.

thought expressed by the proposition negated. This is the only way we *can* define the truth of negations. This is so with all propositions whose truth is defined in terms of the truth of other propositions. They cannot be understood independently of propositions that are involved in the determination of the thoughts they express. A thought expressed by a conjunction is partly determined by thoughts expressed by the conjuncts, a thought expressed by a disjunction partly determined by thoughts expressed by the disjuncts, etc.

Now to the argument. What we are after is a conflict between the Context Principle and the assumption that ‘ \neg ’ in ‘ $\neg Fa$ ’ is a term, given that the thought expressed by ‘ $\neg Fa$ ’ is defined in terms of the thought expressed by ‘ Fa ’. Given the Context Principle, if we have simple signs like ‘ F ’ and ‘ a ’ which could be employed to form a single sentence, then for any one of them to be capable of reference, they must all stand in what Anscombe calls “significant relations” to one another (*IWT*, p. 68). They must, in other words, be taken in the sentence that they form. We may call the “significant relations” between the simple signs that must be in place for the signs to be terms *sentential connections*.

Since negation must convey what is negated, and the thought expressed by a negation is defined in terms of the thought negated, ‘ Fa ’ must both occur in ‘ $\neg Fa$ ’ and express a thought regardless of whether it occurs in ‘ $\neg Fa$ ’. So, ‘ Fa ’ and ‘ $\neg Fa$ ’ must both express a thought, and the second sentence cannot express a thought unless the first one does. If ‘ \neg ’ in ‘ $\neg Fa$ ’ was *not* a term, we could maintain this also the other way around; that ‘ Fa ’ could also not express a thought unless ‘ $\neg Fa$ ’ does. There would then be no separation of the thought expressed by a logically complex proposition and the thoughts expressed by its atomic bases. We find this view in the *Tractatus*. “An elementary proposition,” Wittgenstein says, “really contains all logical operations in itself” (*TLP*, 5.47). However, if ‘ \neg ’ in ‘ $\neg Fa$ ’ was a term, then the thought expressed by ‘ $\neg Fa$ ’ would involve a component not involved in the thought expressed by ‘ Fa ’, namely the object referred to by ‘ \neg ’. The axiom for negations would then define the truth of a negation *partly* in terms of what is negated but would also invoke a logical object. Consequently, if ‘ \neg ’ was a term in a negation, it would have to be an *extra* term, a term that is added to a proposition that already expresses a thought. But the Context Principle does not allow for that.

Suppose that ‘ \neg ’ in ‘ $\neg Fa$ ’ is a term. We can then distinguish between two cases. We could either treat ‘ $\neg Fa$ ’ as made up of three terms: ‘ \neg ’, ‘ F ’, and ‘ a ’, or we could treat it as made up of only two: ‘ \neg ’ and ‘ Fa ’. In the first case, ‘ \neg ’ would stand in a sentential connection

to ‘*F*’ and ‘*a*’ which in turn stand in another sentential connection to each other. In the second case, the sentence ‘*Fa*’ would be treated as a term of a special kind that is itself made up of terms ‘*F*’ and ‘*a*’, and ‘ \neg ’ would only stand in a sentential connection to ‘*Fa*’. In a sense, ‘*F*’ and ‘*a*’ might still be said to occur in ‘ $\neg Fa$ ’ even then, but they would not contribute their referents to the whole but only served to determine some further object—the referent of ‘*Fa*’—that is then contributed to the whole. Which object this is will not make a difference to our argument.

Whichever way we construe the sentence ‘ $\neg Fa$ ’, the conflict with the Context Principle is essentially the same. Given the Context Principle, *all* terms in ‘ $\neg Fa$ ’ must stand in a sentential connection to each other before *any* of them could refer. If ‘ $\neg Fa$ ’ had ‘*F*’ and ‘*a*’ as terms, then neither ‘*F*’ nor ‘*a*’ could refer unless they stand in a sentential connection to ‘ \neg ’ in ‘ $\neg Fa$ ’. But they *must* refer independently of this sentential connection, since the thought expressed by ‘ $\neg Fa$ ’ is defined in terms of the thought expressed by ‘*Fa*’; and so, ‘*Fa*’ must already express a thought and its terms must consequently already refer. If, on the other hand, ‘ $\neg Fa$ ’ was construed as having only two terms, ‘ \neg ’ and ‘*Fa*’, then the new term ‘*Fa*’ could not refer unless *it* stood in a sentential connection to the term ‘ \neg ’. But the referent of ‘*Fa*’, whatever it is, must still be determined compositionally by what the referents of ‘*F*’ and ‘*a*’ are. It may be, for example, that ‘*Fa*’ and ‘*Rbc*’ have the same referent, despite not having all the same terms in the same sentential connections—a candidate for a referent like this would be a truth-value or an intension—but if the propositions do have the same referent, their terms in sentential connections must determine that. Thus, the referent of ‘*Fa*’ would be determined by the thought expressed by ‘*Fa*’. But then, since ‘*Fa*’ must express a thought independently of ‘ $\neg Fa$ ’, ‘*Fa*’ *will* have a referent regardless of its sentential connection to the term ‘ \neg ’. This again conflicts with the Context Principle. So, whichever way we construe ‘ $\neg Fa$ ’, the sign ‘ \neg ’ cannot be a term in it.

To put the argument briefly, if ‘ \neg ’ is a term in a negation, then it stands in a sentential connection to other terms in the negation. For brevity, let us consider only the construal on which negation has only one other term, the negated proposition. In that case, the negated proposition would need to contribute a component to the thought expressed by the negation. Since terms—expressions that contribute components to thoughts—can receive referents only in the context of a proposition, the negated proposition could not have a referent unless considered in the context of the negation. But the thought expressed by the negation is defined

in terms of the thought expressed by the negated proposition; that is, “[o]nly a finished proposition can be negated” (*NB*, p. 25). If ‘ \neg ’ in ‘ $\neg Fa$ ’ is a term, then, by the Context Principle, ‘ Fa ’ taken by itself is not “finished”. But it has to be finished if the truth of ‘ $\neg Fa$ ’ is to be defined in terms of the truth of ‘ Fa ’. Whatever its referent is, it must be determined compositionally by its terms and thus independently of any sentential connection of the negated proposition to the term ‘ \neg ’. And so, the negated proposition *must* have a referent outside the context of the negation. To put the lesson generally, if the signs for logical operations were terms, then Wittgenstein’s point that “[i]t is impossible for words to occur in two different roles: by themselves, and in the proposition” (*TLP*, 2.0122) would carry over to the bases of complex propositions. They could not occur by themselves and as bases if to occur as a base means to be combined with other terms.

The result applies generally to all signs whose semantic value is defined as part of a recursive definition of a thought in terms of other thoughts. For example, if ‘ \wedge ’ in ‘ $p \wedge q$ ’ was a term, then, given the Context Principle, ‘ p ’ and ‘ q ’ could not express thoughts independently of their sentential connection to ‘ \wedge ’, since they (or their terms) could not be assigned objects to be contributed to the thought expressed by the conjunction, unless they (or their terms) were considered as standing in sentential connections to ‘ \wedge ’ in the conjunction. However, which thought is expressed by ‘ $p \wedge q$ ’ is defined in terms of the thoughts expressed by ‘ p ’ and ‘ q ’ and thus depends on ‘ p ’ and ‘ q ’ already expressing thoughts. In parallel with the argument for negations, we may say that only finished propositions could be conjoined. Generality is in no way special in this respect. If ‘ $\forall x$ ’ was a term in ‘ $\forall xA(x)$ ’, then all the instances of the generalization would have to stand in a sentential connection to ‘ $\forall x$ ’ before any one of them (or any terms in them) could refer. But the thought expressed by ‘ $\forall xA(x)$ ’ is defined in terms of the thoughts expressed by its instances. And thus, if ‘ $\forall x$ ’ was a term, the instances of a generalization must express their thoughts independently of the generalization. Given that our argument generalizes in this way, we have demonstrated the Grundgedanke.

The result we have arrived at has some odd consequences. Its scope is much broader than that of the Grundgedanke as we initially formulated it. It covers not only the signs for the standard logical operations but *any* sign whose meaning is given recursively by the truth or falsity of propositions. Whenever we understand a thought expressed by a proposition ‘ q ’ at least partly in terms of the thoughts expressed by other propositions ‘ p_1 ’, ‘ p_2 ’, ..., all signs in ‘ q ’ which are not involved in one of ‘ p_1 ’, ‘ p_2 ’, ... can be shown not to be terms and thus as

having no objects corresponding to them in the thought. At the end of the previous chapter, we considered two schemas in addition to our semantic axioms, one for propositions of the form 'it is surprising that ϕ ' and another for those of the form ' α believes that ϕ '. Given the proposal we considered, the thoughts expressed by these propositions are partly determined by the thought expressed by ϕ . So, the Grundgedanke as we have now defended it covers 'surprising' and the signs α and 'believes' as they occur in ' α believes that ϕ '. No objects can correspond to them when they occur in the propositions in question. This is odd for two reasons.

First, we noted that surprise and belief are things that, on the face of it, are firmly part of the empirical world. The conclusion we have reached does not in fact call this into question, though it requires us to be more careful about what we mean when we speak of such things. When we speak of surprise and belief, we speak of something that *people* have. Empirical investigation of surprise and belief is investigation that concerns the surprised and the believing. Indeed, the same point holds of knowledge. Given the propositional conception of the a priori, we study the propositions known. Given the way of knowing conception, we study the knowers. There is no additional *relation* of knowing between the two. That is why epistemology in which the way of knowing conception is adopted is bound to collapse into cognitive psychology.

The other oddity about the generality of our conclusion is more interesting. In sentences of the form ' α believes that ϕ ', α is a term by stipulation and so refers to an object. Our proof, however, shows that it is not. This is not a paradox. When we introduced propositions of the form ' α believes that ϕ ', we took natural language grammar at face value. It has now turned out that this was a mistake. Sentences of the form ' α believes that ϕ ', if truth is to be defined for sentences, are not possible. Their introduction would require treating the subclause (or the terms in it) as being capable of reference independently of a sentential connection to α and 'believes' *and* as being capable of it only in sentential connection to α and 'believes'. What we have thus shown is that Davidson (1968) was right, at least in broad outline, in treating apparent sentences expressing indirect discourse as composed of two sentences. The proper form of

'Galileo said that the Earth moves',

Davidson suggests, is

'Galileo said that. The Earth moves',

where the first sentence involves a demonstrative to a putative utterance of Galileo's, and the second "gives the content of the subject's saying, but has no logical or semantic connection with the original attribution of a saying" (ibid., p. 106). The direct justification that Davidson gives for the proposal is a bit of empirical evidence suggesting that, in English, the complementary use of 'that' grew out of the demonstrative use. In the Estonian language, a language I am a speaker of, the complementing phrase does not double as a demonstrative. Davidson's proposed evidence, in other words, does not support general conclusions concerning language as such. The real reason for Davidson's proposal is that, once adopted, "[t]he paradox, that sentences (utterances) in *oratio obliqua* do not have the logical consequences they should if truth is to be defined, is resolved" (ibid., p. 108). The paradox that Davidson has in mind is Frege's puzzle that we covered briefly in the previous chapter. Davidson's proposal is an attempt to resolve it without two-level semantics. Of course, one cannot justify a solution to a paradox by the mere fact that it is a solution. We must have some reason for thinking that the solution is right. Given that we are concerned with a definition of truth, that reason must stem solely from an investigation of the nature of truth. We now have, it seems, found a reason like this. The argument for the Grundgedanke does not show that Davidson's positive account about demonstratives being involved in indirect discourse is right. But it does show that the breaking of the apparent sentence into two is required by a definition of truth.

This completes our defence of the Grundgedanke, the idea that the signs for logical operations are not terms and thus that there are no logical objects. Its consequence, as we have seen, is that the truth-values of propositions of logic are not answerable in any way to how things (even logical things) are and thus that propositions of logic do not express thoughts. With this we have shown that logic is indeed "a *totally* different kind than any other science" (*Extracts*, p. 119). In the next and final chapter, we will consider the epistemic implications of this result.

VII

Knowledge of the Laws of Truth

“The meaning of the word “true,” as Frege puts it, “is explained by the laws of truth” (*T*, p. 290). This we noted in chapter V. We are now in a position to make something of it. In short, given the Grundgedanke, the meaning of ‘true’ is not simply explained by logical truths in a way that the occurrence of rain may be explained by temperature and atmospheric pressure. It is explained in the sense that truth is *unpacked* in the laws of truth. In this way, logical knowledge turns out to be both necessary and sufficient for having an understanding of what truth is. An understanding of truth concerns all propositions. To know the meaning of ‘true’ is to know something about all propositions. In this way, as we will see, logical knowledge is presupposed in any recognition of truth and thereby in any acquisition of knowledge. That is why logical knowledge must be a priori.

We have identified a conception of the a priori-a posteriori distinction on which the distinction is primarily one of kinds of truth and derivatively of kinds of justification. Building on remarks by Kant and Berkeley, we then gave an outline of how an account of the apriority of logical truths which would also be an account of logical knowledge might proceed. It might proceed, as we saw, by a defence of Alice’s thesis, that logical falsities cannot be believed. But we left open how the primary distinction between kinds of truth might yield a secondary distinction between kinds of justification. We have now characterized logical truths in enough detail both to defend Alice’s thesis and to spell out the secondary distinction from it.

Building on Dummett’s remarks concerning generalizations, we will first cover what it is to have an understanding of truth. As we will see, it involves having a general grasp of the range of possibilities specified by truth-evaluability. This will be our way of spelling out the idea of epistemic possibilities that we introduced in chapter IV. Assuming the Grundgedanke that we defended in chapter VI, we will then show that this range of possibilities is the range of logical possibilities. From that, we will demonstrate Alice’s thesis and develop an account of logical knowledge. We will then turn to the secondary distinction between kinds of justification and show how it follows from our account. Finally, we will address an odd consequence of the account, that logical knowledge cannot be acquired. We will take inspiration from Plato concerning how to square this with appearances.

1. An Understanding of Truth

An understanding of truth is a general understanding. It is an understanding of what it is for any proposition to be true, or equivalently, of what it is for any sentence to express a thought. This is assumed when Convention T is taken as a criterion of adequacy for a definition of truth. To accept Convention T as a criterion is to assume that truth is characterized only by stating for each sentence what it is for *it* to be true, and hence that truth cannot be explained by any ‘one-liner’.

To understand a proposition—to know what it is for it to be true—one must have an understanding of what truth is. That is, to discern the thought expressed by a *given* proposition, one must already have a general understanding of what it is for *any* proposition to be true. An understanding of a single proposition thus presupposes an understanding that, in some sense, concerns them all. Our first task is to explain how this could be.

A parallel with generalizations suggests itself. To understand a generalization is to know what it is for each of its instances to be true. As we have noted in previous chapters, an understanding of this sort cannot be had by virtue of understanding the instances separately. One can only understand a generalization by having a *general* understanding of its instances. We have seen what this general understanding consists in. As Frege explains it, the instances of a generalization are specified by their form. An understanding of a generalization involves an understanding of the instances solely *as* propositions of the specified form. This sort of general understanding is open-ended. A form does not specify which exemplars it has. By the same token, it leaves open which thoughts are expressed by the exemplars. An understanding of a generalization consequently does not involve an understanding of each instance of the generalization taken separately. But the appeal to form is needed for more than just to account for the possibility of understanding generalizations in one go. Dummett explains Frege’s account thus.

We understand the universally quantified statement because we have, as it were, a *general* grasp of the totality which constitutes the domain of quantification—we, as it were, survey it in thought as a whole . . . We do not need, in any sense, to know what objects there are which belong to that domain: we need only to have a determinate conception of what it is to belong to the domain, so that it is objectively determined, for each object, whether it is in that domain or not. (*PoL*, pp. 517–518)

To appreciate Dummett's point, we need to say more about domains of quantification. For each object in the domain, a generalization has an instance that concerns that object. In our definition of truth, we avoided an explicit mention of a domain by letting the domain be given by the terms in the meta-language. The reason for this, though we could not state it at the time, is that, given the Context Principle, we can only identify objects in thoughts. By letting the domain be given by terms in meta-language propositions which are used to assign thoughts to object-language sentences we always construe objects as constituents of thoughts. In this way, objects are, so to speak, never taken out of thoughts. Given the Context Principle, talk of elements of a domain is a shorthand way of speaking of thoughts which have the elements as constituents. The relevant thoughts are those expressed by the instances of the generalization.

Dummett's point that we need to have "a determinate conception of what it is to belong to the domain, so that it is objectively determined, for each object, whether it is in that domain or not" (ibid.) thus amount to the following. One does not understand a generalization by virtue of understanding its instances, as we have already noted, but rather by knowing what it is for any proposition to *be* its instance, that is, by knowing what it is for a proposition to be of the form specified in the generalization. An understanding with a general reach of this kind ensures that for a proposition that one might come across only later, or not at all, it is *already* "objectively determined" whether *it* is an instance of the generalization and thus whether its truth or falsity bears on the truth or falsity of the generalization. Only then will a generalization have a truth-value by virtue of what we apprehend when we understand it, that is, by virtue of the thought that it expresses. The truth or falsity of a generalization must, so to speak, be determined *in advance* of an understanding of any given instance of it. For illustration, suppose one understands ' $\forall xA(x)$ ', and then later finds that ' $A(c)$ ' is false. The recognition of the falsity of ' $A(c)$ ' must itself be sufficient for the recognition of the falsity of ' $\forall xA(x)$ '. If a further thought was required—such as the thought that ' $A(c)$ ' is an instance of ' $\forall xA(x)$ '—then a general understanding of the instances of a generalization would not be sufficient for knowing what it is for the generalization to be true.

As with the instances of a generalization, whatever it is that we know in having an understanding of truth must, in parallel with Dummett's point, objectively determine all propositions without requiring that we know individually which thoughts are expressed by them. Only then can our understanding of truth capture what is required for each proposition to have a truth-value regardless of whether the thought it expresses is apprehended. An understanding of truth, in other words, must be open-ended in the same way as an

understanding of generalizations needs to be. In our definition of truth, this is reflected in all propositions being specified by form alone, just as the instances of a generalization are. By this we construe an understanding of truth as presupposing knowledge of what it is for something to be of a specified form. To understand what it is for something to be true one must know what it is for something to be of any form specified in an adequate definition of truth. Only then can an understanding of truth be open-ended in the required way.⁵³

The analogy with generalizations cannot, however, be taken too far. We have also seen that it should be possible to understand a generalization from a single instance. There is no room for a parallel point in the case of an understanding of truth. As we noted above, to have an understanding of a given proposition, one must *already* have a general understanding of truth. One cannot first understand one proposition and then learn from that what truth is in the way one could gain an understanding of a generalization from its instance. Where the analogy with generalizations breaks down, in other words, is in how an understanding of truth is to be acquired. An understanding of truth, unlike an understanding of a generalization, cannot be gained from any propositions. This, of course, invites the question how an understanding of truth *can* be gained. But an answer to that question does not matter for our purposes. What matters is whether this understanding, which must precede the recognition of anything as truth-evaluable, might already be sufficient for knowledge of some truths. If it is, then this knowledge could not have been gained in the way we gain knowledge by experience. Experience tells us that something or other is the case. It thus *presupposes* an understanding of truth. If there is knowledge that one already has by virtue of an understanding of truth, then one has it independently of experience.

To pursue the idea of knowledge by virtue of an understanding of truth, we need a framework in which the existence of such knowledge might be decided. We sketched an outline for a framework like this in chapter IV. Back then, we introduced an intuitive notion of an epistemic possibility and defined a corresponding notion of believability in terms of it. In the next section, we will replace the intuitive notion with something we can specify precisely.

⁵³ A consequence of an understanding of truth always going via knowledge of what it is for something to be of a specified form is that a proposition is always to be recognized as an exemplar of a form. To recognize a specific proposition is to recognize a form specific enough to have only one exemplar. This is essentially the same point that we noted already in our defence of the Context Principle: one cannot identify the proposition by its terms but must instead identify the terms in the proposition. That is why terms cannot receive referents outside of a proposition.

2. Truth-Possibilities and Logical Space

The nature of truth determines what it is to have a truth-value. In doing so, it gives us a range of possibilities, the broadest range of possibilities that there is. To treat ‘ p ’ as possibly true, in whatever sense of ‘possible’, we treat ‘ p ’ as affording the value true. We may thus consider the range of possibilities that is determined solely by truth-evaluability. Adapting terminology of Wittgenstein’s (*TLP*, 4.3), we may call these possibilities *truth*-possibilities. We may see truth-possibilities as relating to a definition of truth in a way that is similar to how the instances of a generalization relate to the generalization. A definition of truth specifies truth-possibilities by their form. And we may construe an understanding of truth as a grasp of the range of all truth-possibilities.⁵⁴

A belief need not be true, but it must have a truth-value. To believe something is to attribute truth to it. Thus, what can be believed can also be true in the broadest sense of ‘can’. That is, all beliefs are truth-possibilities. But we may also define the broadest sense of ‘can be believed’ in terms of truth-possibilities. Since to believe something is to attribute truth to it, we may introduce a sense of ‘can be believed’ in which something can be believed just in case it affords the value true; that is, just in case it is a truth-possibility. This is a non-psychological sense of ‘can be believed’, a sense of the sort we sketched in chapter IV. From now on, it is how we will read Alice’s thesis, the claim that logical falsities cannot be believed. Since truth-possibilities are determined by the nature of truth, a definition of truth by itself lets us decide whether something can be believed in this sense. We thus have a criterion by which to assess Alice’s thesis.

The range of truth-possibilities is determined by the range of thoughts, the range of things about which it can be asked ‘is it true?’. We can construe truth-possibilities with the help of the contemporary notion of a possible world if we adjust the notion slightly. In the literature on semantics and modal logic, worlds are often treated as primitives of the theory. But this is a mistake. T-sentences do not involve reference to worlds. They are not of the form

(T^w) ‘ p ’ is true in a world w if and only if p in w

⁵⁴ The meaning of ‘grasp’ is both intuitive and unclear and is thus liable to lead to a false sense of clarity. To grasp a totality, as we will use the term, is to know what it is for something to belong to the totality. This matches Dummett’s use of the term to mean knowledge of what it is to belong in the domain of quantification. However, whenever we speak of grasping, this can be spelled out in terms of knowledge of what it is for something to be of some *form* or other where the form is always a propositional form.

but of the form

(T) ‘ p ’ is true if and only if p

It is only if worlds are themselves defined in terms of thoughts that we can say that a proposition is true in a world without violating Convention T. The allusion to worlds in (T^w) will then not be a switch to a different meaning of ‘true’ but only a way of considering attributions of truth to specific propositions in the broader context of a definition of truth. The appeal to worlds may then always be dropped to yield a regular T-sentence. But having the notion of a world will simplify the argumentation to come. A grasp of the range of all truth-possibilities can then equally be construed as a grasp of the range of all worlds, where worlds are understood in terms of truth-possibilities.

In the previous chapter, we defended Wittgenstein’s Grundgedanke, that there are no logical objects. Building on it, we may now consider what does and does not belong in the range of all worlds and thus what can and cannot be believed. An implication of the Grundgedanke is that the recursive clauses in our definition of truth must be read in a certain way. As Wittgenstein puts it,

The sense of a truth-function of p is a function of the sense of p . (*TLP*, 5.2341)

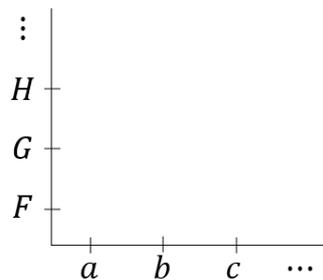
That is, a thought expressed by a complex proposition like ‘ $p \wedge q$ ’, if it expresses a thought, is a function of the thoughts expressed by its immediate bases ‘ p ’ and ‘ q ’. The same applies to the bases. In this way, a thought expressed by any complex proposition is a function of the thoughts expressed by its *atomic* bases. Given the Grundgedanke, no additional component beyond what is involved in thoughts expressed by atomic propositions is introduced by the recursive clauses. This implies that we must define worlds—where each proposition is true or false—in terms thoughts expressed by atomic propositions.

We can construe each atomic proposition—an expression of a state of affairs—as determining a position in a coordinate system with objects referred to in the atomic proposition serving as the coordinates. The coordinate system is what Wittgenstein calls *logical space*, the objects the *logical coordinates*, and any set of points in logical space determined by atomic propositions a *logical place*.

A proposition determines a place in logical space. The existence of this logical place is guaranteed by the mere existence of the constituents—by the existence of the proposition with a sense. (*TLP*, 3.4)

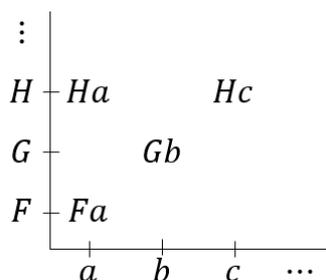
The propositional sign with logical co-ordinates—that is the logical place. (3.41)

For the sake of clarity, we will reserve the term ‘logical place’ for positions determined by atomic propositions only. A complex proposition with different atomic bases may then be said to determine a set of logical places. The idea of logical space given by objects as coordinates is illustrated by Sullivan (2001, p. 95) (who credits Roger White for the idea). For example, atomic propositions that contain two terms can be seen as giving a two-dimensional coordinate system with one dimension for relations and another for individuals like so:



This is a way of writing the states of affairs expressed by ‘*Fa*’, ‘*Ga*’, ‘*Gb*’, etc. Each form of atomic propositions gives a coordinate system like this. We may call them atomic coordinate systems. For an atomic proposition to have two terms as constituents just is for the coordinate system to have two dimensions. For an atomic proposition to have three terms as a constituent is for the co-ordinate system to have three dimensions, etc. All the atomic coordinate systems together form the logical space. They are linked by virtue of shared dimensions. For example, ‘*Fa*’ and ‘*Rab*’ each determine a logical place in different atomic coordinate systems, but both coordinate systems share at least the dimension on which the object *a* is a coordinate.

An atomic proposition can receive a truth-value, and nothing in the definition of truth requires it to receive one truth-value over another. A logical place determined by the atomic proposition is thus a truth-possibility. As Wittgenstein puts it, “[i]n geometry and logic alike a place is a possibility: something can exist in it” (*TLP*, 3.411). What can exist in a logical place is a state of affairs. If an atomic proposition is true, the state of affairs obtains, filling its logical place. And if the atomic proposition is false, the state of affairs does not obtain, leaving the place empty. So, for each distribution of truth-values to all atomic propositions, we get a different distribution of states of affairs in logical space. In the case of atomic propositions with two terms, the distribution may yield the following:



We may now define worlds in the way Wittgenstein does.

If all true elementary propositions are given, the result is a complete description of the world. The world is completely described by giving all elementary propositions, and adding which of them are true and which false. (*TLP*, 4.26)

The world is all that is the case. (1)

What is the case—a fact—is the existence of states of affairs. (2)

In this sense of ‘world’, any distribution of truth-values to all atomic propositions determines a world. A world is a distribution of states of affairs throughout the whole of logical space. In truth-table notation, if we consider a truth-table with a column for each atomic proposition, then worlds may be seen as determined by the rows. A proposition expresses a truth-possibility just in case the thought it expresses is true in some world.

An objection that this construal of possibility may invite is that, although there is a state of affairs for every atomic proposition, there might not be an atomic proposition for every state of affairs.⁵⁵ If there isn’t, then our construal of truth-possibilities fails to capture the truth of all thoughts. The objection is closely related to the worry that we addressed in chapter V regarding nameless objects in an account of generality; and our response to it is also similar. We have already seen that a definition of truth—which is also a definition of proposition—defines atomic propositions in general terms only. It does not settle which atomic propositions there are but only what semantic role they play: an atomic proposition is an expression of a state of affairs. It is this general conception that our account of truth-possibilities invokes when it appeals to the truth-values of all atomic propositions. Our account thereby covers all atomic propositions in an open-ended range and will thus apply equally to atomic propositions that are yet to be formulated.

⁵⁵ I am grateful to my examiner José Zalabardo for bringing this objection to my attention.

The outline of the framework we have arrived at is familiar from contemporary semantics and modal logic. All and only the logical possibilities are truth-possibilities. Logical truths are true in all worlds, and logical falsities are true in none. What makes our framework different is that the range of worlds we have arrived at is given solely in terms of a definition of truth. It thus corresponds to the broadest sense of possibility there is. Whatever affords the value true is true in at least one world, even if that truth is in some more restricted sense impossible. And whatever is not true in any world does not afford the value true. To say of a logical falsity that it is true in a world would be to change the meaning of ‘true’ to something not captured by Convention T. This explains why there are no logically impossible worlds, worlds where logical falsities are true.⁵⁶ The nature of truth itself excludes them. From this, Alice’s thesis follows.

3. Alice’s Thesis and Logical Knowledge

We saw previously that whatever can be believed, whatever a subject can attribute truth to, is a truth-possibility. We have now seen that every truth-possibility is a logical possibility. There are thus no attributions of truth to logical falsities and thus no belief in logical falsities. Whenever it appears to us that a subject believes a logical falsity, it is a confusion on *our* part concerning the meaning of the word ‘true’, or, which amounts to the same, a confusion concerning which proposition it is that the subject believes to be true.

In chapter IV, we gave a rough sketch of a distinction between two kinds of knowledge implied by Alice’s thesis, knowledge that involves thoughts and knowledge that does not. But to spell this out, we first need to say more about logical truths as they are construed in our newly developed framework.

Logical truths are true in all worlds but in a special way. As we saw in the previous chapter, given the Grundgedanke, logical truths are intrinsically true. Their truth is in any case not answerable to which states of affairs obtain, but, given the absence of logical objects, this amounts to their truth not being answerable at all to how things are. Logical truths, as Wittgenstein puts it, are “unconditionally true” (*TLP*, 4.461). Since a definition of truth is a

⁵⁶ Logically impossible worlds are introduced, for example, by Jago (2014) as well as by many others. For a detailed overview and defence of the idea, see Berto & Jago 2019. Generally, a defence of the idea of impossible worlds goes via a rejection of the Grundgedanke. If it is assumed that logical relations are components of thoughts, there is nothing to stop us from considering worlds where those relations hold in logically impossible ways.

definition of proposition, we then arrived at the result that logical truths are true solely by virtue of which propositions they are. We now have a more precise way of stating this which lets us clarify the relationship between logical truths and thoughts.

The base clause of a definition of truth—by defining atomic propositions—determines a logical space which is enough for all propositions that express thoughts to be truth-evaluable. For the truth of a logical truth, it does not matter how states of affairs are distributed in logical space. A logical truth will be true regardless. But the logical space itself does matter. For a logical truth to be true its bases must be truth-evaluable. For a disjunction to be true at least one of its disjuncts must be. The logical truth ' $p \vee \neg p$ ' is in no way special in this respect. What is special about it is only that, unlike with disjunctions normally, the truth-*evaluability* of its bases is already enough to ensure its truth. Nothing needs to be the case for its truth. Moreover, given the absence of logical objects, nothing *beyond* the truth-evaluability of its bases is required. For the proposition ' $p \vee \neg p$ ' to be true just is for there to be a logical space. This holds generally of all logical truths.

So, given the Grundgedanke, for a logical truth to be true just is for there to be a logical space, a coordinate system of all worlds. In a definition of truth, the coordinate system itself is given only schematically. We do not need to state which atomic propositions there are and thus which coordinates or dimensions the logical space has. The schematic stipulation is enough to define other propositions recursively as functions of atomic propositions which in turn is already enough for logical truths to be true. So, one can know of any logical truth that it is true without needing to know anything about what belongs to the totality of all worlds and thus without needing to know which thoughts are expressed by which propositions. Still, for a logical truth to be true there must *be* thoughts expressed by propositions. There must still be logical coordinates (objects referred to by terms in atomic propositions) which determine a logical space. On the other hand, since for a logical truth to be true just is for there to be a logical space, knowledge of logical truths is all that is required for a grasp of the totality of worlds and thus for knowledge of what it is for anything to be true. The understanding that logical knowledge affords of truth is thus general only. It leaves open all questions concerning specific thoughts. All it settles is that there are thoughts.

As we saw in chapter IV, given Alice's thesis, our logical beliefs, if they are knowledge, must be a priori knowledge. Since logical truths are not expressions of thought, no amount of

investigation—empirical or otherwise—into how things are can be relevant to learning of a logical truth that it is true. But we can now give a more detailed argument to this effect.

We have just seen that one has logical knowledge just in case one has a general understanding of truth. Previously, we saw that to discern a *given* thought one must already have a general understanding of truth. We may thus conclude that to discern a thought, one must already have logical knowledge. The conclusion is not merely that logical knowledge is required for non-logical *knowledge*. It is that one must have logical knowledge already to *understand* a proposition, regardless of whether one takes the proposition to be true. By knowing logical truths, we discern the possibility of truth of all propositions, logical and non-logical. On the other hand, non-logical truths alone, since they cannot be understood without a prior understanding of truth, cannot help us discern the possible truth of any proposition, including their own. We *have* logical knowledge because we *can* discern thoughts. Since experience tells us what is or is not the case—since it tells us which thoughts are true and which are not—logical knowledge must come before experience in just the sense that Plato, Kant, and others had in mind. It must be a priori knowledge.

The argument has some similarities with BonJour's (1998) argument for the existence of a priori knowledge. Positing that beliefs justified by direct experience are "particular rather than general in their content" (ibid., p. 4), BonJour notes that it should be possible to infer from them "in a way that brings with it epistemic justification" to beliefs "whose content goes beyond direct experience or observation" (ibid.). Otherwise, "the upshot is a quite deep form of scepticism" (ibid.). So, without a priori knowledge many of our other beliefs, for example, "beliefs about the past, the future, and the unobserved aspects of the present" (ibid.) would not be knowledge at all. This argument, if sound, would only be enough to show that our epistemic situation would be dire without a priori knowledge. It would not show that we *have* a priori knowledge.

Our conclusion is different. Logical knowledge, we have shown, is presupposed not just by non-logical knowledge but by the recognition of the possibility of truth and falsity. This is because to have logical knowledge is to have a *general* understanding of what it is for anything to be true. We have logical knowledge because we are, in a word, thinkers. The difference between having and not having logical knowledge, we might say, is the difference between having and not having the capacity for thought. The absence of logical knowledge, in Kant's

word, “annihilates” not just knowledge but any capacity for thought (*CPR*, A151).⁵⁷ Still, we have so far simply assumed that our logical beliefs are knowledge. All we have really shown is that, *if* they are knowledge, they are a priori knowledge. We have yet to explain why those beliefs should be knowledge in the first place.

4. Two Kinds of Knowledge Revisited

We saw in chapter IV that if logical truths do not express thoughts, we must distinguish between two kinds of knowledge. By now, we have demonstrated this distinction. Non-logical knowledge is knowledge of matters of fact. Logical knowledge is not. Knowledge that it is either raining or not is of a different kind than knowledge that it is raining. Back in chapter IV, we noted that this is a first step towards an epistemologically deep distinction between a priori and a posteriori knowledge. It is only a first step because, as stated, it does not yet pertain to justification. It is a direct consequence of our primary distinction between kinds of truth. In this section, we will see how the primary distinction also yields a secondary distinction between kinds of justification.

The result that logical truths do not express thoughts was a consequence of there being no logical objects. An issue that this gives rise to is that between two distinct logical truths, for example, ‘ $Fa \vee \neg Fa$ ’ and ‘ $\forall x \exists y (Rxy \leftrightarrow Ryx)$ ’, there is no difference other than a syntactic difference. Not only does neither express a thought but the signs for logical operations involved in them also introduce no other distinction between them beyond the syntactic distinction. So, although there are many logical truths, there is no difference in the truth of one of them and the truth of another. For one logical truth to be true is for all of them to be true. Since we have arrived at this consequence while operating with a *fine-grained* notion of a thought, a further consequence of it is that there is only one thing to know in logic. To know one logical truth is to know them all, and to fail to know one is to fail to know them all. There are, of course, syntactic differences between logical truths, but syntactic differences do not distinguish cases of knowledge. Although we do often speak of knowledge of the truth of a proposition out of convenience, what we then mean is still knowledge of *what is the case*. The primary sense in which, in having knowledge, the subject recognizes truth is expressed by the truth-operator,

⁵⁷ Kant’s view is slightly different from what we have argued for. As we saw previously, Kant maintains only that it is belief in logical falsity that annihilates belief as such. We have shown that the absence of belief in logical truth already does the trick.

not the truth-predicate. If it was the truth-predicate—if knowledge that p was knowledge that the *proposition* ‘ p ’ is true—then all knowledge would concern language. We might introduce a notion of knowledge like this to speak of logical knowledge, thereby taking advantage of the syntactic distinctions between logical truths. But a new notion does not do away with the old. Moreover, to make sense of the new meta-linguistic notion of knowledge, we would still need to make sense of it in terms of the old notion, as we will see later.

A justified belief is a belief that, if true, is knowledge. Questions concerning justification are non-trivial only when there is a difference between knowledge and mere true belief. For there to be this difference, false belief must be possible. If one can have a false belief, there should be some assurance that what one believes is not false. David Lewis (1986) has proposed that this point by itself implies that all beliefs in necessary truths thereby have a special status.

[I]f it is a necessary truth that so-and-so, then believing that so-and-so is an infallible method of being right. If what I believe is a necessary truth, then there is no possibility of being wrong. That is so whatever the subject matter of the necessary truth and no matter how it came to be believed. (Lewis 1986, pp. 114–115)

A belief in a necessary truth cannot be false and thus, Lewis suggests, one is guaranteed to be justified in believing it. The proper response to this, given by Schechter (2010) and Williamson (2013) among others, is that the possibility of a belief being false in this context should not be understood as concerning the possible falsity of what one already believes. What is at issue is rather the possibility of forming a false belief in similar circumstances or in the face of the same evidence. This being a test of the reliability of the belief-forming mechanisms or the strength of the evidence. However, if it was impossible not to have a false belief in place of the true one, then a point analogous to Lewis’s would hold.

What we noted about the inseparability of cases of logical knowledge carries over to belief. To fail to believe one logical truth is to fail to believe them all. As we saw previously, to believe a logical truth is to have a general understanding of truth. Thus, to fail to believe even one logical truth is to fail to have this general understanding. But without a general understanding of *what it is* for anything to be true, one could not recognize anything as truth-evaluable and thus could not form any beliefs. In particular, one could then not form false beliefs. That is, one could then not be wrong. The possibility of error, in other words, implies

that one already believes every logical truth.⁵⁸ Thus, there cannot be any need for assurance that one's logical beliefs are true. In this way, the distinction between knowledge and belief collapses in logic. We may say that, in the case of logical truths, belief implies justification. This gives us a distinction between two kinds of justification. There is justification for which some epistemic work, say, empirical investigation is called for, and there is justification for which there isn't. Logical beliefs are vacuously justified and thereby count as knowledge. And whatever we take justification in a posteriori knowledge to require, *that* is not required for logical knowledge. So, logical beliefs are knowledge but not a posteriori knowledge. They are thus a priori knowledge.

Admittedly, our standard notions of belief, knowledge, and justification produce odd results when it comes to logic. Much of our argumentation up to this point is reminiscent of doing arithmetic with 0. But just as 0 is still a number, logical knowledge is still knowledge. In having logical knowledge, one has a justified true belief. The odd results concerning justification and belief are consequences of what *truth* amounts to in the case of logical truths. The truth of a logical truth is not answerable to how things are. There is thus no thought whose truth it is. Our notions of belief and knowledge are more naturally applied in connection to thoughts. But so long as it is meaningful to speak of truth, they *do* apply. It should be stressed that in speaking of truth we are still concerned with our *intuitive* conception of truth that is captured by Convention T. To be ours, the conception itself must be intuitive. Its consequences need not be.

We should not be satisfied with establishing a *distinction* between a priori and a posteriori knowledge if this does not also help us clarify how the two kinds of knowledge are related to each other. The distinction should have a point. Given our argument for the apriority of logical knowledge, its purpose in relation to non-logical knowledge is straightforward. We have noted that in the absence of logical knowledge one cannot come to have false beliefs, since to form *any* belief, one must already have a general understanding of truth. For the same reason, one cannot, in the absence of logical knowledge, come to have logically true beliefs. In other words, it is not possible to *learn* logical truths. One must always already know them. But in *having* logical knowledge, one can come to know whatever else. It is in this way that logical knowledge—which is a priori—makes a posteriori knowledge possible. It provides the space for there to be a posteriori knowledge. Knowledge that we must have, in order to be thinkers

⁵⁸ For this reason, logical knowledge is also immune to scepticism. Sceptical doubt—the possibility of widespread error—cannot be raised unless logical knowledge is already granted.

at all, makes it possible for us to be ignorant of other matters and to overcome that ignorance by gaining new knowledge. My coffee mug does not know what the weather will be like tomorrow. In this respect, I and the coffee mug have something in common. The difference between us is that I can *come* to know about tomorrow's weather whereas the coffee mug cannot. And that difference, we have seen, is the difference between having logical knowledge and not having it.

5. Anamnesis

There is a loose end to be tied up concerning logical knowledge. It may be put in this way: we obviously *do* learn logical truths; it *takes an effort* to recognize something as logically true; and, absent this effort, we can have logically true beliefs that fall short of knowledge. This, in effect, is the White Queen's response to Alice. We are now ready to address it.

It is certainly right that we are aware of some logical truths and not others, and that the recognition of something as logically true takes an effort. Our current conception of knowledge does not account for that. Nor does it help to introduce some other propositional attitude, like awareness or recognition or rational insight (cf. BonJour). The reasoning that runs for belief or knowledge that $p \vee \neg p$ runs equally for awareness or recognition that $p \vee \neg p$ as well as for any other propositional attitude characterizable with the help of a logical truth.

The answer to what it is that we recognize, with effort, as logically true cannot be a thought. But nor is it supposed to be. What we recognize as logically true is a proposition. We noted previously that there are still syntactic differences between logical truths. Before we take advantage of them, however, it should be noted that Wittgenstein, whose ideas we have been heavily building on, does not himself think this. As he says, "[i]f p follows from q and q from p , then they are one and the same proposition" (*TLP*, 5.141). The reason for this is something we too adopted previously, namely the distinction between accidental and essential features of propositions. Essential features of propositions, Wittgenstein says, "are those without which the proposition could not express its sense" (*TLP*, 3.34). Consequently, "what is essential in a proposition is what all propositions that can express the same sense have in common" (*TLP*, 3.341). Since logically equivalent propositions express the same sense, Wittgenstein concludes that they are the same proposition. But then all logical truths come out as the same proposition as well. These identity conditions for propositions are the result of a careless characterization

of what essential features of propositions are. Although it is correct to say that essential features are those without which the proposition could not express its sense, it is correct to say it because they are features by virtue of which the proposition *does* express its sense. Propositions are individuated in terms of their form to which semantic significance is assigned in a definition of truth. We cannot say that ' $p \vee \neg p$ ' and ' $\forall x \exists y (Rxy \leftrightarrow Ryx)$ ' are of the same form because there must be a distinction between the forms ' $\phi \vee \psi$ ' and ' $\forall v \phi(v)$ '.

As we noted previously, given the syntactic distinctions between logical truths, we can introduce a notion of knowledge where syntactic distinctions matter. But before we do, we should be clear about what this further notion of knowledge is for. Its *sole* purpose is to account for appearances. The White Queen's reason for maintaining that we can believe logical falsities, fail to have logical knowledge, etc. comes down to this: it *looks* that way, and no argument from the nature of truth will make that appearance go away. This is also the starting assumption in much of contemporary investigation of a priori knowledge and thus cannot be dismissed out of hand.

We may schematically define our new notion of knowledge, which we may call knowledge*, in terms of knowledge as follows:

S knows* that p iff S knows that ' p ' is true.

In a sense, knowledge* is linguistic knowledge. It concerns propositions and the semantic values of the expressions that make up the proposition. The sort of idea that the notion of knowledge* captures has been proposed, for example, by Stalnaker (1984) and more recently by Rayo (2013). Rayo gives a helpful illustration.

Suppose you are asked 'Are there any tapirs in the area?'. In order to answer this question, you need two pieces of information. First, you need some non-linguistic information: you need to know whether there are any tapirs in the area. Second, you need some linguistic information: in particular, you need to know what 'tapir' means. (Rayo 2013, p. 108)

To know* whether there are tapirs in the area, as we have defined the notion, you need to know what 'tapir' means. If you already know (in the normal sense) whether there are tapirs in the area and are told what 'tapir' means, you will be able answer the question correctly without more ado. That is the sense in which knowledge* that p depends on knowledge that p .

Propositions, we have seen, have essential and accidental features. The same proposition can be written in different ways. But since any given proposition does have accidental

features—a proposition is always spelled out in one way or another—those accidental features still bear on what a subject has knowledge* of. Whether a subject knows* that there are tapirs in the area depends on whether the subject knows which proposition is articulated as ‘there are tapirs in the area’. Thus, unlike with logical knowledge, we can lack and gain logical knowledge* since we can fail to recognize the proposition in the words and are capable of overcoming this failure; and there is seemingly room for unjustified logically true belief*—where belief* that p is belief that ‘ p ’ is true—since we can guess of some given words that they are a formulation of a logical truth and be right by coincidence.

At this point, it may be wondered why we shouldn’t simply say that contemporary accounts of logical knowledge and perhaps of a priori knowledge more generally really concern what we are now calling knowledge* and that our odd results concerning logical knowledge therefore have little theoretical significance. But, as we noted in connection to Rayo’s illustration, we can only have knowledge* if we have corresponding knowledge proper. The reason is reflected in T-sentences, sentences of the form

‘ p ’ is true if and only if p .

What is known in knowing that p is captured by the right-hand side of a T-sentence of ‘ p ’, and what is known* is captured by the left-hand side. In our definition of truth, the truth of the left-hand side of a T-sentence is defined in terms of the truth of the right-hand side. So, there is nothing that could be known* independently of corresponding regular knowledge. Since our definition of truth concerns propositions in general terms only, one does still need additional linguistic knowledge about the specific proposition ‘ p ’ to have knowledge* that p , but this is only because one needs to identify the proposition to have a *belief* about it. Everything that matters for the *truth* of the proposition—and therefore for the truth of the belief that the proposition is true—is captured by the right-hand side of the proposition’s T-sentence; that is, by what one knows in the normal sense of ‘know’. The answer to why a true belief* that p is knowledge* is the same as the answer to why a true belief that p is knowledge. Any language-related cognitive effort needed for knowledge*—since it plays no role in regular knowledge—can only play a merely enabling role in knowledge*.

Since the non-linguistic component of logical knowledge* is vacuous, any cognitive effort required to gain logical knowledge* boils down solely to the study of language and is thus all merely enabling. All that one *can* do to gain logical knowledge* is to figure out whether a given sentence is a formulation of a logically true proposition. That is what we do, for

example, when we consult the definition of truth in producing a semantic proof. The difficulty of coming to know* that $p \vee \neg p$ is precisely the difficulty of figuring out the truth of the sentence ' $p \vee \neg p$ ' from the definition of truth. Because we already have logical knowledge proper, we know that the deductive steps we take in the process are valid. Having identified two propositions, we know their conjunction to be true; having identified a generality, we know its instances to be true, etc. In this way, we rely on our logical knowledge in identifying the propositions which express it. And this can obviously be done poorly or not at all. So, the apparent possibility of unjustified logically true belief*—which was among the White Queen's worries—is to be regarded as the possibility of unjustified belief concerning which proposition is spelled out in which words; something which has nothing to do with logic.

Our account of the apparent learnability of logical truths in terms of knowledge* closely resembles Plato's account of the apparent learnability of truths of form with which we began our investigation; that what appear to be cases of learning are really cases of *anamnesis*, a kind of recollection. The prompts that lead to *anamnesis*—like the questions that Socrates asks—are obviously insufficient to explain why the person knows the truth of what they thereby come to think. The exchange in *Meno* between Socrates and the slave boy is meant to illustrate just this. The goal is to show that the boy can “find the knowledge within himself” (85d). Just as we introduced the notion of knowledge*, Plato too oscillates between two notions of knowledge. For example, Socrates also says: “[T]he man who does *not know* has within himself true opinions” (85c, emphasis added). Plato attributes knowledge in this second sense based on how the subject responds to questions, or more generally, on whether the subject assents to propositions. Knowledge within the person is thus mapped to propositions that the subject assents to which Plato takes to constitute knowledge in another sense. But the reason why knowledge in the second sense is still knowledge in Plato's account, why it has the epistemic status that it does have, is due to the underlying knowledge in the first sense that has been inside the person all along.

Given the account of logical knowledge that we have arrived at, it is only by something like *anamnesis* that the appearance of gaining logical knowledge *can* be explained. The investigation of the nature of logical knowledge commonly stems from a puzzlement regarding a specific practice of operating with propositions. On what basis or with what right do we make claims that we infer from other claims, for example, or criticise claims on the basis of consistency? What our findings show is that the answers to those questions regarding the legitimacy of the practice do not lie *within* this practice. They lie in the nature of truth which

makes the practice possible. It is also why one can indeed teach logic to anyone capable of thought solely by asking the right questions. For the truth, as Socrates puts it, “is always in our soul” (86b). Our difficulty “is only an—enormous—difficulty of expression” (*NB*, p. 40).

Conclusion

Logical truths, I have argued, can be known independently of experience. In other words, logical knowledge, knowledge of logical truths, is a priori knowledge. But, as stated, there are two ways of taking this claim: either as a statement about subjects and their cognitive abilities or as a statement about the nature of logical truths. To the two readings correspond two conceptions of the a priori-a posteriori distinction. My first task was to distinguish these conceptions clearly from each other.

Reading the claim in the first way, we construe the a priori-a posteriori distinction primarily as a distinction between two ways of knowing: an experiential and a non-experiential way. At least since Saul Kripke's influential work, this conception of the distinction has been predominant in contemporary epistemology. But we saw that it has not always been so and found reasons to be sceptical of its prospects in epistemology. If we read the claim in the second way, as concerning the nature of logical truths, it amounts to a statement that logical truths are of the sort that knowledge of them does not call for experience. The a priori-a posteriori distinction is then construed as a distinction between kinds of truth, between truths the knowledge of which calls for experience and truths the knowledge of which does not. I argued that this conception, the propositional conception, can yield an epistemically significant a priori-a posteriori distinction of the sort that the way of knowing conception simply cannot.

It is characteristic of the propositional conception that the apriority of knowledge is solely a matter of the nature of what is known. With this in mind, I proceeded to clarify the nature of logical truths, the truths one knows in having logical knowledge. Given the propositional conception, it is only an investigation of this kind that could show logical knowledge to be a priori. Logical truths, as Frege puts it, are propositions that express the laws of truth. That is one respect in which logical truths differ from all other propositions. They are the only propositions whose truth is determined by what truth is. To clarify the nature of logical truths one must therefore investigate the nature of truth itself. This I did by following the standard approach popularized by Tarski's work, adjusting it where necessary by insights from Frege and Ramsey.

The nature of truth, as Tarski has argued, is to be captured in a recursive definition that implies for each proposition its corresponding T-sentence where the thought expressed by the proposition is given. To take a definition like this to reflect the nature of logical truths is, in

Frege's words, to give "pride of place" to truth in logic. There can be nothing in logic that is not understood ultimately in terms of truth. This implies Frege's famous Context Principle, that it is only in the context of a proposition that a term can have an object as its referent, since objects—the referents of terms in propositions—must likewise be understood in terms of truth. I argued that this in turn implies what Wittgenstein calls his fundamental idea, or *Grundgedanke*, that the signs for logical operations cannot be construed as having any sort of objects corresponding to them and thus as contributing components to thoughts. I showed that a consequence of this is that logical truths are propositions that are true but do not express thoughts, or *sinnlos*, as Wittgenstein describes them. Logically true propositions turn out instead to have their truth-values intrinsically, that is, solely by virtue of which propositions they are. It is this insight about the nature of logical truths that I relied on to then argue that logical knowledge is a priori.

The possibility of thinking of something as true presupposes that one has a general understanding of what truth is. To clarify how this general understanding is to be construed, I relied on Dummett's remarks on what an understanding of a quantified proposition consists in. In having an understanding of a quantified proposition, one has a general grasp of its instances. Likewise, in having an understanding of truth, one has a general grasp of a range of possibilities, a range delineated by what truth itself is. I argued that, since logical truths do not themselves express thoughts, one has this general grasp just in case one knows every logical truth. An understanding of truth alone is thus already sufficient for having all the logical knowledge there is to have. Since one needs to have an understanding of truth to acquire knowledge by experience (or indeed by any other way), logical knowledge turns out to be presupposed by a posteriori knowledge and must itself be a priori.

The finding that there is this relationship between logical knowledge and a posteriori knowledge is perhaps more significant than the demonstration of the apriority of logical knowledge itself. To have logical knowledge is to have an understanding of what truth is. In this way, logical knowledge makes it possible for us to discern truth and falsity. Without it we could have no grasp of thoughts, that is, of possible situations, and thus could not acquire *any* knowledge. This is the role that logical knowledge plays in our lives as knowers. Logical knowledge is what makes that life possible.

Bibliography

- ALLEN, REGINALD E. "Anamnesis in Plato's "Meno and Phaedo"," *The Review of Metaphysics* 13(1): 165–174.
- ANSCOMBE, G. E. M. 1959. *An Introduction to Wittgenstein's Tractatus*. St. Augustine's Press.
- ARISTOTLE 2016, *Metaphysics*, C. D. C. Reeve (transl.). Hackett Publishing Company.
- ARISTOTLE 1935, *On the Soul (De anima)*. In Aristotle, *On the Soul, Parva Naturalia, On Breath*, W. S. Hett (transl.). William Heinemann Ltd, Harvard University Press. pp. 8–203.
- ARMSTRONG, DAVID M. 2004. *Truth and Truthmakers*. Cambridge University Press.
- AYER, ALFRED J. 1952. *Language, Truth and Logic*. Dover Publications.
- AYER, ALFRED J. 1956. *The Problem of Knowledge*. Macmillan.
- AYER, ALFRED J. (ed.) 1959. *Logical Positivism*. The Free Press.
- BEALER, GEORGE 2002. "Modal Epistemology and the Rationalist Renaissance," in *Conceivability and Possibility*, T. S. Gendler, J. Hawthorne (eds.). Clarendon Press. pp. 71–125.
- BERGER, ALAN 2011. "Kripke on the Incoherency of Adopting a Logic," in *Saul Kripke*, A. Berger (ed.). Cambridge University Press. pp. 177–207.
- BERKELEY, GEORGE 1998. *A Treatise Concerning the Principles of Human Knowledge*, J. Dancy (ed.). Oxford University Press.
- BERTO, FRANZ, MARK JAGO 2019. *Impossible Worlds*. Oxford University Press.
- BIRKHOFF, GARRETT, JOHN VON NEUMANN, 1936. "The Logic of Quantum Mechanics," *Annals of Mathematics*, 37(4): 823–843.
- BONJOUR, LAURENCE 1998. *In Defense of Pure Reason: A Rationalist Account of A Priori Justification*. Cambridge University Press.
- BONJOUR, LAURENCE 2001. "Review: Précis of *In Defence of Pure Reason*," *Philosophy and Phenomenological Research* 63(3): 625–631.
- BOSTOCK, DAVID 1988. "Necessary Truth and A Priori Truth," *Mind* 97(387): 343–379.
- BURGE, TYLER 2000. "Frege on Apriority," in *New Essays on the A Priori*, P. Boghossian, C. Peacocke (eds.). Clarendon Press. pp. 11–42.
- CARNAP, RUDOLF 1932. "The Elimination of Metaphysics Through Logical Analysis of Language," A. Pap (transl.), in AYER 1959, pp. 60–81.

- CARROLL, LEWIS 1998. *Through the Looking-Glass, and What Alice Found There*. In L. Carroll, *Alice's Adventures in Wonderland and Through the Looking-Glass*. Edited by H. Haughton. Penguin Books.
- CASULLO, ALBERT 2003. *A Priori Justification*. Oxford University Press.
- CASULLO, ALBERT 2015. "Four Challenges to the A Priori-A Posteriori Distinction," *Synthese* 192: 2701–2724.
- CASULLO, ALBERT *forthcoming*. "A Defense of the Significance of the A Priori-A Posteriori Distinction," in *The A Priori: Its Significance, Grounds, and Extend*, D. Dodd, E. Zardini (eds.). Oxford University Press. URL = <https://www.unl.edu/philosophy/Casullo_Defense-OUP-Final.pdf>
- CHISHOLM, RODERICK M. 1957. *Perceiving: A Philosophical Study*. Cornell University Press.
- CONNOR, EARL, RICHARD FELDMAN 1985. "Evidentialism," *Philosophical Studies* 48(1): 15–34.
- CRESSWELL, MAX J. 1975. "Hyperintensional Logic," *Studia Logica: An International Journal for Symbolic Logic* 34(1): 25–38.
- DAVIDSON, DONALD 1967. "Truth and Meaning," in DAVIDSON 2001, pp. 17–36.
- DAVIDSON, DONALD 1968. "On Saying That," in DAVIDSON 2001, pp. 93–108
- DAVIDSON, DONALD 1973. "In Defence of Convention T," in DAVIDSON 2001, pp. 65–75.
- DAVIDSON, DONALD 2001. *Inquiries into Truth and Interpretation*, Second Edition. Clarendon Press.
- DAVIDSON, DONALD 1982. "Empirical Content," in D. Davidson 2001. *Subjective, Intersubjective, Objective*. Clarendon Press. pp. 159–175.
- DESCARTES, RENÉ 2006. *Meditations, Objections, and Replies*, R. Ariew, D. Cress (transl., eds.). Hackett Publishing Company.
- DEVITT, MICHAEL 2005. "There Is No *a Priori*," in *Contemporary Debates in Epistemology*, M. Steup, E. Sosa (eds.). Blackwell Publishing. pp. 105–115.
- DUMMETT, MICHAEL 1973. *Frege: Philosophy of Language*. Duckworth.
- DUMMETT, MICHAEL 1983. "The Philosophical Basis of Intuitionistic Logic," in *Philosophy of Mathematics: Selected Readings*, Second Edition, P. Benacerraf, H. Putnam (eds.). Cambridge University Press. pp. 97–129.
- FODOR, JERRY 2008. *LOT 2: The Language of Thought Revisited*. Oxford University Press.
- FREGE, GOTTLIB 1956. "The Thought: A Logical Inquiry," *Mind* 65(259): 289–311.
- FREGE, GOTTLIB 1960. "Function and Concept," P. Geach (transl.), in G. Frege, *Translations from the Philosophical Writings of Gottlob Frege*, Second Edition, P. Geach, M. Black (eds.). Basil Blackwell. pp. 21–41.

- FREGE, GOTTLÖB 1964. *The Basic Laws of Arithmetic: Exposition of the System*, M. Furth (transl., ed.). University of California Press.
- FREGE, GOTTLÖB 1969. *Nachgelassene Schriften*, H. Hermes, F. Kambartel, F. Kaulbach (eds.). Felix Meiner.
- FREGE, GOTTLÖB 1979. *Posthumous Writings*. H. Hermes, F. Kambartel, F. Kaulbach (eds.), P. Long, R. White (transl.). University of Chicago Press.
- FREGE, GOTTLÖB 1980. *The Foundations of Arithmetic*, Second Edition, J. L. Austin (transl.). Northwestern University Press.
- GEACH, PETER 1965. "Assertion," *The Philosophical Review* 74(4): 449–465.
- GETTIER, EDMUND 1963. "Is Justified True Belief Knowledge?" *Analysis* 23(6): 121–123.
- GOLDMAN, ALVIN 1979. "What Is Justified Belief?" in *Justification and Knowledge: New Studies in Epistemology*, G. S. Pappas (ed.). D. Reidel Publishing Company. pp. 1–23.
- GOWERS, TIMOTHY, MICHAEL NEILSON 2009. "Massively Collaborative Mathematics," *Nature* 461: 879–881.
- GRICE, PAUL H., PETER STRAWSON 1956. "In Defense of a Dogma," *The Philosophical Review* 65(2): 141–158.
- HEIM, IRENE, ANGELIKA KRATZER 1998. *Semantics in Generative Grammar*. Blackwell Publishers.
- HEMPEL, CARL G. 1950. "The Empiricist Criterion of Meaning," in AYER 1959, pp. 108–129.
- JAGO, MARK 2014. *The Impossible: An Essay on Hyperintensionality*. Oxford University Press.
- KANT, IMMANUEL 2007. *Critique of Pure Reason*. M. Weigelt (transl., ed.), based translation by M. Müller. Penguin Books.
- KITCHER, PHILIP 1980. "A Priori Knowledge," *The Philosophical Review* 89(1): 3–23.
- KRIPKE, SAUL 1980. *Naming and Necessity*. Basil Blackwell.
- LEIBNIZ, GOTTFRIED W. 1989. *Philosophical Essays*, R. Ariew, D. Garber (eds., transl.). Hackett Publishing Company.
- LEIBNIZ, GOTTFRIED W. 1996. *New Essays on Human Understanding*, P. Remnant, J. Bennett (eds., transl.). Cambridge University Press.
- LEWIS, DAVID 1986. *On the Plurality of Worlds*. Blackwell Publishing.
- MARCUS, RUTH BARCAN 1983. "Rationality and Believing the Impossible," *The Journal of Philosophy* 80(6): 321–338.
- PADRÓ, ROMINA 2015. *What the Tortoise Said to Kripke: the Adoption Problem and the Epistemology of Logic*. CUNY Academic Works, URL = <http://academicworks.cuny.edu/gc_etds/603>

- PERRY, JOHN 2001. *Reference and Reflexivity*. CSLI Publications.
- PLATO 2002. *Five Dialogues*, Second Edition, G. M. A. Grube (transl.), J. M. Cooper (revis.). Hackett Publishing Company, Inc.
- *Meno*, pp. 58–92.
 - *Phaedo*, pp. 93–154.
- PRIEST, GRAHAM 2006. *In Contradiction: A Study of the Transconsistent*, expanded edition. Oxford University Press.
- PUTNAM, HILARY 1975. “The Logic of Quantum Mechanics,” in P. Putnam, *Mathematics, Matter and Method: Philosophical Papers*, Volume 1. Cambridge University Press. pp. 174–197.
- PUTNAM, HILARY 1983. “‘Two Dogmas’ Revisited,” in P. Putnam, *Realism and Reason: Philosophical Papers*, Volume 3. Cambridge University Press. pp. 87–97.
- QUINE, WILLARD VAN ORMAN 1936. “Truth by Convention,” in *Readings in Philosophical Analysis* 1949, H. Feigl, W. Sellars (eds.). Appleton-Century-Crofts, Inc. pp. 250–273.
- QUINE, WILLARD VAN ORMAN 1951. “Two Dogmas of Empiricism,” *The Philosophical Review*, 60(1): 20–43.
- QUINE, WILLARD VAN ORMAN 1986. *Philosophy of Logic*, Second Edition. Harvard University Press.
- QUINE, WILLARD VAN ORMAN, J. S. ULLIAN 1978. *Web of Belief*, Second Edition. McGraw-Hill, Inc.
- RAMSEY, FRANK P. 1925. “Universals,” in RAMSEY 1931, pp. 112–134.
- RAMSEY, FRANK P. 1927. “Facts and Propositions,” in RAMSEY 1931, pp. 138–155.
- RAMSEY, FRANK P. 1929. “Last Papers,” in RAMSEY 1931, pp. 212–269.
- RAMSEY, FRANK P. 1931. *The Foundations of Mathematics and other Logical Essays*, R. B. Braithwaite (ed.). Routledge and Kegan Paul Ltd.
- RAYO, AGUSTIN 2013. *The Construction of Logical Space*. Oxford University Press.
- RECANATI, FRANÇOIS 2012. *Mental Files*. Oxford University Press.
- SCHECHTER, JOSHUA 2010. “The Reliability Challenge and the Epistemology of Logic,” *Philosophical Perspective* 24: 437–464.
- SCHLICK, MORITZ 1932/33. “Positivism and Realism,” D. Rynin (transl.), in AYER 1959, pp. 82–107.
- SLUGA, HANS 2002. “Frege on the Indefinability of Truth,” in E. H. Reck (ed.), *From Frege to Wittgenstein: Perspectives on Early Analytic Philosophy*. Oxford University Press.
- SMILEY, TIMOTHY 1963. “Relative Necessity,” *The Journal of Symbolic Logic* 28(2): pp. 113–134.
- SULLIVAN, PETER 2001. “A Version of the Picture Theory,” in *Ludwig Wittgenstein, Tractatus Logico-Philosophicus*, W. Vossenkuhl (ed.). Akademie Verlag. pp. 89–110.

- STALNAKER, ROBERT 1984. *Inquiry*. MIT Press.
- TARSKI, ALFRED 1944. “The Semantic Conception of Truth: and the Foundations of Semantics,” *Philosophy and Phenomenological Research* 4(3): 341–376.
- TARSKI, ALFRED 1956. “The Concept of Truth in Formalized Languages,” in A. Tarski, *Logic, Semantics, Metamathematics: Papers from 1923 to 1938* (second edition), J. H. Woodger (transl.), J. Corcoran (ed.), Hackett Publishing Company. pp. 152–278.
- TRAVIS, CHARLES 2018. “What Structure Lurks in the Minds of Men? Frege vs Fodor,” in *The Philosophy of Charles Travis: Language, Thought, and Perception*, J. Collins, T. Dobler (eds.). Oxford University Press. pp. 50–90.
- WILLIAMSON, TIMOTHY 2000. *Knowledge and Its Limits*, Oxford University Press.
- WILLIAMSON, TIMOTHY 2007. *Philosophy of Philosophy*, Blackwell Publishing.
- WILLIAMSON, TIMOTHY 2013. “How Deep is the Distinction between A Priori and A posteriori Knowledge?” in *The A Priori in Philosophy*, A. Casullo, J. Thurow (eds.). Oxford University Press. pp. 291–312.
- WITTGENSTEIN, LUDWIG 1979. *Notebooks 1914–1916*, Second Edition, G. H. von Wright, G. E. M. Anscombe (eds.), G. E. M. Anscombe (transl.). The University of Chicago Press.
- *Notes on Logic, 1913*, Appendix I, pp. 93–107.
 - *Extracts from Wittgenstein’s Letters to Russell, 1912–20*, Appendix III, pp. 120–132.
- WITTGENSTEIN, LUDWIG 1922. *Tractatus Logico-Philosophicus / Logisch-Philosophische Abhandlung*, Side-by-Side Edition, Version 0.59 (12 May 2021), C. K. Ogden (transl.), D. F. Pears, B. F. McGuinness (transl.), K. C. Klement (ed.), URL = <<http://people.umass.edu/klement/tlp/>>
- WRIGHT, CRISPIN 2004. “Intuition, Entitlement and the Epistemology of Logical Laws,” *Dialectica* 58(1): 155–175.
- YALCIN, SETH 2014. “Semantics and Metasemantics in the Context of Generative Grammar,” in *Metasemantics: New Essays on the Foundations of Meaning*, A. Burgess, B. Sherman (eds.). Oxford University Press. pp. 17–54.
- ZAGZEBSKI, LINDA 1994. “The Inescapability of Gettier Problems,” *The Philosophical Quarterly* 44(174): 65–73.

Abbreviations

ANSCOMBE

IWT — *Introduction to Wittgenstein's Tractatus*

DUMMETT

PoL — *Frege: Philosophy of Language*

FREGE

FA — *Foundations of Arithmetic*

PW — *Posthumous Writings*

T — "The Thought: A Logical Inquiry"

KANT

CPR — *Critique of Pure Reason*

TARSKI

SCT — "The Semantic Conception of Truth"

WITTGENSTEIN

NB — *Notebooks 1914–1916*

NL — *Notes on Logic*

TLP — *Tractatus Logico-Philosophicus*