Introduction

Proof-theoretic semantics is a theory of how the meanings of logical constants, expressions like 'not', 'and', 'or', 'if-then', 'all' and 'some', are determined by the rules of inference governing them. Its origins are found in Gentzen's work. In his systems of natural deduction, each logical constant is governed by introduction rules that specify under which conditions a formula with that constant as the main operator may be derived and elimination rules that specify what may be derived from such a formula. Gentzen noticed a 'remarkable systematic' in the 'inference patterns' for the logical constants and proposed what might be called *Gentzen's Thesis*:[1]

> The introductions constitute, so to speak, the "definitions" of the symbols concerned, and the eliminations are in the end only consequences thereof, which could be expressed thus: In the elimination of a symbol, the formula in question, whose outer symbol it concerns, may only "be used as that which it means on the basis of the introduction of this symbol". An example may elucidate what is meant: the formula $\mathfrak{A} \supset \mathfrak{B}$ can be introduced if there is a deduction of $\mathfrak{B}$ from the assumption formula $\mathfrak{A}$. Should one now want to use it again in the elimination of the symbol $\supset$ [...], then this can be done precisely by deducing $\mathfrak{B}$ immediately from a proof of $\mathfrak{A}$, because $\mathfrak{A} \supset \mathfrak{B}$ records the existence of a deduction of $\mathfrak{B}$ from $\mathfrak{A}$. Note that in this there is no need to take into account an "intuitive sense" of $\supset$. (Gentzen, 1934: 189)

Gentzen's Thesis has received thorough philosophical treatment and transformation into a comprehensive theory in Michael Dummett's and Dag Prawitz's work. Their views provide much of the background to proof-theoretic semantics.

This book studies a problem proof-theoretic semantics faces in relation to what may reasonably be claimed to be the most important logical constant: *negation*. I argue that the meaning of negation cannot be defined within Dummett's and Prawitz's framework. I defend a solution that is based on the observation that their theory only appeals to a primitive notion of *truth*, tied to notions of verification, justification, grounds or warrants of assertions, which is preserved by inferences. I argue that it also

[1] My translation. See Gentzen (1969: 80f) for the standard translation.

needs to appeal to a corresponding *negative primitive*, namely a notion of *falsity*, tied to notions of falsification or defeat of assertions. On the way to this conclusion, I consider and reject various alternative attempts to solve the problem of negation, supported by metaphysics, philosophical logic and philosophy of language.

The first chapter of this book situates proof-theoretic semantics within philosophical logic and the philosophy of language. The core idea is that the meanings of the expressions of a language are determined by their use. Applied to the logical constants, it is the view that the meanings of the logical constants are determined by their use in deductive arguments, which is represented in Gentzen's systems of natural deduction. Much of the chapter focuses on Dummett's account of a theory of meaning as a systematic description and rational reconstruction of speakers' capacities, the knowledge that enables them to use a language to communicate. Dummett's views are complex and multi-layered, so I restrict myself to an impressionist exposition of material that provides the philosophical foundations of the following chapters.

Moving from language to logic, the second chapter expounds proof-theoretic semantics. Considerations within the philosophy of language rehearsed in Chapter 1 motivate constraints on the form of rules of inference. Gentzen's Thesis is spelled out as Dummett's and Prawitz's proposal that the meaning of a logical constant may be defined by its introduction and elimination rules, if these are *in harmony* or, more precisely, *stable*. Intuitively, for Dummett and Prawitz harmony means that in the course of a deduction the elimination rules for a logical constant allow the retrieval from a formula with the constant as main operator of no more than what one is entitled to retrieve given the grounds for its assertion as specified by the introduction rules. Stability obtains if the converse is also the case and the introduction rules demand as grounds for the assertion of a formula no more than what is required given the consequences of its assertion as specified by the elimination rules. If stability obtains between the introduction and elimination rules of a logical constant, the grounds and consequences of a formula containing it as main operator balance each other perfectly. According to Dummett, the grounds and consequences of any assertion must balance each other, but it is particularly fruitful to apply this idea to the logical constants, as there it can be cashed out in terms of formal properties of rules of inference and deductions. I develop the general framework of rules to be used in the remainder of the book, a formally precise account of harmony and stability. I present a method for how to read off introduction from elimination rules and the other way round. I define notions of harmony and stability in terms of this method that differ slightly from Dummett's and Prawitz's notions, but, I argue, capture their intention more fruitfully than other proposals. For instance, the rules for **S4** necessity are harmonious according to my definition, as they are of a certain form to be specified in Chapter 2: from $\Box A$ infer A, and from A infer $\Box A$ if all the premises on which A depends are of the form $\Box B$. They are not, however, stable, according to my definition, due to the restriction imposed on the application of the introduction rule. Contrast with the rules for $\supset$: from A and $A \supset B$ infer B, and given a deduction of B from A, infer $A \supset B$, discharging A. These rules are stable as they have the right form and do not impose restrictions on their application. The capacity to apply the rules for $\Box$ requires a previous understanding of $\Box$, as the constant is referred to in the conditions of the application of the introduction rule: to apply the rule a speaker needs to understand the meaning of premises of the form $\Box B$. Hence the meaning of $\Box$ cannot be given entirely by those rules. By contrast, nothing but the capacity to follow rules of inference is needed to acquire an understanding of the meaning of $\supset$. Thus, I argue, the meanings of logical constants governed by stable rules are given purely by rules of inference, whereas to specify the meanings of those constants the rules of which are harmonious but not stable, something else is needed in addition. On this account of harmony and stability, the rules for intuitionist logic are all stable, those of classical logic are not, as predicted by Dummett and Prawitz. The chapter ends with a formal section proving results due to Prawitz on the normalisation of natural deduction, which will be useful for reference in later chapters of this book.

Chapter 3 is devoted to the problem of negation within proof-theoretic semantics. In a nutshell, it is the following. $\neg A$ is defined as $A \supset \bot$. The stable rules for $\supset$ are conditional proof and *modus ponens*. The stable rules for $\bot$ are *ex falso quodlibet* and no introduction rule. $\bot$ is supposed to be the ultimate absurdity. It can never be asserted, as it has no introduction rule, and entails everything. Consider, however, the case of a language in which all atomic sentences are independent of each other, just as they are treated in formal logic, and they all happen to be true. Then $\bot$ can be true. Inferring an atomic proposition from a true $\bot$ would not lead from truth to falsity. Neither would inferring a sentence composed of $\supset$, $\vee$ and $\wedge$, as they are all true, if the atomic sentences they are composed of are true. Both A and $A \supset \bot$ can then be true. In a language with only independent atomic sentences and logical constants governed by stable rules of inference, negation would not mean what one would normally expect it to mean. More precisely, the rules governing $\bot$ do not guarantee that it is the ultimate absurdity and always false, as, in admittedly arcane circumstances, it can be true. The rules governing ⊥, however, were supposed to exclude this possibility. Thus they do not impose the intended meaning on ⊥, as they do not guarantee that ⊥ is always false.

The intuitionist proof-theoretic semanticist objects to the classicist that the meaning of classical negation cannot be given purely by rules of inference. The problem of negation, however, shows that the classicist and the intuitionist are in the same boat. The meaning of intuitionist negation cannot be given purely by rules of inference either. Neither position can give a satisfactory account of the meaning of negation. Negation is obviously a very important logical constant. It features in principles such as the law of excluded middle and those codifying the relation between truth and falsity: $\neg A$ is true iff A is false. Formal systems without negation lack the philosophical interest of systems with negation. Any theory about the logical constants needs to say something about negation.

The problem of how to define negation within proof-theoretic semantics has been discussed before, but never in as much detail as here. Chapter 3 traces the complex nature of Dummett's arguments against classical negation and only in this context can the full force of the problem be appreciated. Proof-theoretic semantics can be seen as imposing restrictions on legitimate rules of inference. It aims to single out one logic as the correct one. The claim is that only the negation of intuitionist, but not classical logic, can be explained entirely in terms of the rules of inference governing it, and hence that intuitionist logic is the proof-theoretically justified one, while classical logic is not. Chapter 3 presents various arguments that the logic justified by proof-theoretic semantics is intuitionist and that classical logic is defective. For each argument, I give a corresponding classicist response. The responses, however, have a drawback: they culminate in the argument that the meaning of negation cannot be defined in the framework presented in Chapters 1 and 2. Thus, the final argument to defend classical logic rests on a fundamental shortcoming of the entire approach of proof-theoretic semantics. If abandoning proof-theoretic semantics is not the desired option, it is mandatory to investigate how it can be modified so as to give a satisfactory account of negation. This is the aim of the remaining chapters of the book: the quest for a satisfactory modification that remains within the spirit, if not the letter, of the system of proof-theoretic semantics set out in Chapters 1 and 2.

The key to my approach is an analysis of the primitives of proof-theoretic semantics. As already indicated, Dummett's and Prawitz's theory relies only on positive primitives in addition to rules of inference: affirmation, assertion and truth. A major insight of the path that the book takes is that *negative* as well as positive primitives are required at the foundations of the theory. I consider three options of extending proof-theoretic semantics by negative primitives: to accept that negation is an undefinable primitive, to add a primitive speech act of denial, and to add a notion of falsity. Before tackling those, I consider a fourth option that is popular with some proof-theoretic semanticists and constitutes an attempt to stay within the paradigm of admitting only positive primitives: to add a primitive notion of incompatibility.

Defining negation in terms of a primitive notion of incompatibility is common within the wider context of rule-based accounts of meaning, such as inferentialism. Robert Brandom and Christopher Peacocke argue that the negation of p is its 'minimal incompatible': it is the proposition incompatible with p and entailed by all propositions incompatible with p. Neil Tennant argues that $\neg p$ can be derived if p entails mutually incompatible propositions. I argue that Brandom and Peacocke cannot guarantee that the minimal incompatible of a proposition always exists, and that Tennant's account suffers from the complementary problem, namely to show that there is only one negation of a proposition. The two approaches can be fruitfully combined to solve both problems. However, I argue that the resulting view only justifies minimal logic, which is too weak to provide a satisfactory account of negation, or, at best, Tennant's idiosyncratic intuitionist relevant logic. It is preferable to look elsewhere for an amendment of proof-theoretic semantics.

The approach of explaining negation in terms of incompatibility is also popular amongst metaphysicians without any stake in proof-theory: how can there be negative truths if everything that exists, and thus every truth maker, is positive? This mirrors the restriction to only positive primitives I diagnose in proof-theoretic semantics. I argue, however, that even independently of proof-theoretic semantics, incompatibility is not a good choice of primitive. There are no good metaphysical reasons to accept that there is such a primitive relation, and from an epistemological perspective, negation is understood much better than incompatibility. The discussion of negative truths within metaphysics is not taken up in much detail within proof-theoretic semantics, but despite the negative conclusion of Chapter 4, there is some promise for the unification of two disparate fields of philosophical inquiry.

A slightly defeatist-sounding approach to solving the problem negation poses to proof-theoretic semantics is to accept that the meaning of negation cannot be defined in terms of rules of inference, and to add negation itself as a further primitive to the theory. Accepting something as a primitive is not to give up. The undefinability of negation by rules of inference in Dummett's and Prawitz's framework is a good reason to adopt the approach. There are also independent reasons to accept that negation is undefinable: every theory needs primitives, and negation is such a basic concept that it looks like an ideal choice for a primitive. It may also be possible to say a bit more about why that is so. Peter Geach, influenced by Frege, suggested that to understand a concept, a speaker needs to understand its negation, too, so that the meaning of a concept is inseparably tied to the meaning of its negation. In Chapter 5 I argue, however, that there is no satisfactory way of implementing such an approach in proof-theory. It would demand that the other logical constants get their meaning not just from rules of inference for them, but also from rules for their interaction with negation. I argue that no additional rules can add anything to the meaning of the constants as given by the usual rules, and so any further rules are superfluous, and hence there is no way of implementing this strategy of dealing with the problem of negation within proof-theoretic semantics. Quite independently of being an unsatisfactory approach within proof-theoretic semantics, Geach's view relies on a primitive notion of predicate negation. I argue for Frege's view that there is only sentential negation and that apparent examples of predicate negation should be analysed in terms of sentential negation.

Chapter 6 investigates the option of adding a primitive speech act of denial. This approach has become popular with proof-theoretic semanticists in the wake of the development of logics for assertion and denial. *Bilateralism* aims to specify the meanings of expressions not only in terms of the grounds and consequences of asserting them, but also in terms of the grounds and consequences of denying them. The idea stems from Huw Price, who developed it in order to answer Dummett's challenge of providing a satisfactory use-based theory of sense that justifies classical logic but shows intuitionist logic to be defective. Ian Rumfitt developed bilateral logics for assertion and denial and argues that in these systems only the rules for classical but not those for intuitionist logic are in harmony (or, more precisely, stable). I show that Rumfitt's claim is wrong by formalising an intuitionist bilateral logic in which all logical constants are governed by stable rules of inference. A close examination of Price's argument for classical logic shows that he is not entitled to his conclusion either, and that, in fact, the principles he employs in his account may well lie better with intuitionist logic. Price and Rumfitt agree that, for methodological reasons, bilateralism is to be preferred over unilateralism only if it justifies classical logic while ruling out intuitionist logic. As I show that intuitionist logic can be justified on Rumfitt's account and that Price's bilateralist argument for classical logic is unsuccessful and that accepting intuitionist logic may even be advantageous for his account, I conclude that the bilateralist approach has no methodological advantage over the simpler, more straightforward unilateralist approach and should therefore be abandoned.

In the final chapter, I defend and develop the option of adding a notion of falsity as an additional primitive. My approach incorporates insights discussed in earlier chapters. Like Geach and the bilateralists, I agree that there are two aspects of meaning: meaning is determined not only by truth conditions, but also by falsity conditions. I also take on an idea found in Tennant's writings: it is essential for the learnability of language that the truth-values of some propositions change. For all I know, however, a disembodied Cartesian mind need not ever experience primitive incompatibilities and instead spend its time deducing logical, mathematical or other *a priori* truths. A better way of putting Tennant's point is that even if a such a mind never asserts a falsehood, it would need to grasp what it is for a sentence to be false. It is essential to understanding language that a speaker grasps truth and falsity. This builds on an insight of Frege's: 'A false thought must, although not as true, nonetheless sometimes be acknowledged as indispensable: first, as the sense of an interrogative sentence, secondly as a component of a hypothetical compound thought, and thirdly in negation'[2] (Frege, 1918: 147). From the proof-theoretic perspective, false sentences are essential to *reductio ad absurdum*.

My approach can be differentiated from truth-conditional semantics and stays within the spirit of proof-theoretic semantics, as stability of rules of inference is still the crucial aspect in determining the meanings of the logical constants. There are three options of how to proceed with the justification of deduction. One is to go the intuitionist route and to claim that to solve the problem of negation, it suffices to introduce a primitive notion of falsity to explain the meaning of $\bot$, while the rest, in particular the justification of logical laws within the standard systems of natural deduction, can stay the same. However, if truth as well as falsity are primitives, this should be mirrored in the proof system: ordinary natural deduction only has truth-preserving rules, so on the present account, there should also be rules appealing to the notion of falsity. Natural deduction can be modified so as to incorporate not only truth-preserving but also falsity-preserving rules of inference. One such system has been proposed

[2] My translation. See Geach and Black (1952: 122) for the standard translation.

by Heinrich Wansing. But Wansing's system has no structural rules codifying relations between truth- and falsity-preserving deductions. This creates a certain imbalance in a principle of stability which I formulate for his calculus. The book ends with a presentation of the formal component of my solution to the problem of falsity. Following Wansing, in addition to the common truth-preserving rules of inference of Gentzen's systems of natural deduction, I add falsity-preserving rules. I also introduce structural rules that allow what I call the fusion of truth- and falsity-preserving deductions. These rules are of a kind that is new to the literature. They are effectively a restricted version of Cut in systems with multiple conclusions. The resulting system takes on some of the deductive power of multiple conclusion logics, while staying within the framework of natural deduction, thus evading objections to multiple conclusion logics that have been given within proof-theoretic semantics. The system only appeals to conceptual resources also appealed to by natural deduction, apart, of course, from an additional appeal to a notion of falsity. The resulting logic is classical, but satisfies all the criteria of proof-theoretic semantics. I prove a theorem corresponding to normalisation for natural deduction, which entails that proofs in normal form have the subformula property. This establishes the philosophical adequacy of my system.

The chapter closes with reflections on how truth and falsity and their place in a theory of meaning can be elucidated by Bernhard Weiss's observation that linguistic practice is governed not only by principles concerning the correctness of assertions, but also by principles governing the conditions under which assertions must be retracted. Dummett argues that the notion of the correctness of assertions is the source of the concept of truth. The falsity of an assertion is the prime reason for retracting it. Thus I would suggest that the notion of the retraction of assertions is the source of the concept of falsity.