

Chapter 2

Modality and Logic

2.1 History and Beyond

Within the tradition of analytic philosophy and modern logic C.I. Lewis [99, 100] was probably the first to systematically investigate possible formalizations of modal notions and Lewis opted for an operator treatment.¹ He proposed calculi for a non truth-functional two-place sentential operator ‘ \rightarrow ’ which he proposed to read as *strict implication*. That is, for two sentences ϕ, ψ the expression $\lceil \phi \rightarrow \psi \rceil$ forms a new sentence of the language. A one place modal operator ‘ \Box ’ which can be read as ‘necessarily’ could be defined in the calculus as follows

$$(\star) \quad \Box \phi := (\top \rightarrow \phi)$$

Around the mid thirties of the 20th century Lewis, Lukasiewicz, Becker and others had already developed a wide range of modal calculi which proved useful in analyzing modal notions in natural language and mathematics. Von Wright’s “Essay on Modal Logic” [176] in 1951 probably marks the beginnings of the applications of these modal calculi to a broader class of modalities, namely epistemic and deontic modalities.

¹Hugh MacColl [102] already as early as 1906 introduced a two-place, non truth functional operator for implication. However, he did not provide a rigorous axiomatic treatment like Lewis nor did he develop a proper semantic analysis of this notion. See Goldfarb [50] for more on this and an essay on the origins and the development of modern modal logic. Rahman [137] gives an overview and interpretation of MacColl’s work.

Before we continue two disclaimer seem appropriate at this point. First, of course, modalities and modal logic have been discussed long before the rise of analytic philosophy and modern logic. We refer the interested reader to the classical Kneale and Kneale [85].

More generally, the introductory historical remarks do not purport to be scholarly in any relevant sense but should introduce the reader to the debate between predicate and operator conceptions of modality. For historical remarks on the development of modern modal logic from the perspective of modal operator logic valuable information may be found in Goldfarb [50] and Blackburn et al. [15], Chap. 1.

The operator approach received a further boost by the work of Hintikka [73, 74], Kanger [79], and Kripke [89, 90] who developed the so-called possible world semantics for modal operator logic which provided a rigorous semantic interpretation of the modal notions under consideration. This semantics which can be considered as a model-theoretic rendering of Leibniz' possible world picture of modalities was widely considered as highly intuitive and as the correct analysis of modal notions. At least since the development of this semantics modal operator logic and possible world semantics is at the heart of formal philosophizing.

However, according to Carnap [26] and Quine [128, 130, 131] Lewis was confused in treating the modal notions, and especially the notion of strict implication, as operators rather than as predicates. Lewis took his operator ' \rightarrow ' to represent the notion of logical implication which was meant to be understood in terms of logical consequence or in terms of derivability. None of these notions does, as Lewis had correctly noticed, coincide with the notion of material implication. By introducing the operator for logical implication he thus meant to distinguish between the uses of 'imply' by Russell and Whitehead in their *Principia Mathematica*. Unfortunately, Lewis failed to realize that both logical consequence and derivability as they are commonly understood are relations between sentences and thus no matter for which notion we opt strict or logical implication ought to be formalized by a predicate rather than an operator. From a syntactic point of view logical implication should thus be applied to names of sentences rather than the sentences themselves. But then the modality defined by the appropriate version of (\star), i.e.,

$$(\star\star) \quad Nt := \rightarrow ('T', t)$$

is not an operator but a predicate. So if necessity is understood as definable from the notion of logical implication and logical implication is understood as either logical consequence or derivability, then necessity will be a predicate and not an operator.² Moreover, if we stick to this analysis and, additionally, allow ourselves to the distinction between object- and metalanguage, necessity should be understood as an expression of the metalanguage.

It is unclear whether Lewis in his early writings was in fact aware of the distinction between operators and predicates, as from a historic perspective the distinction between use and mention only became clear during the 1930s subsequent to the work of Gödel, Tarski, and, especially, Carnap and Quine. But if no distinction is made between using sentences and mentioning them, then the distinction between operators and predicates of sentences will evidently be blurred.

Whereas the foregoing establishes that Lewis was confused about use and mention this by no means shows that it is confused to conceive of the modal notions of natural language as operators. We need not accept the above definition of necessity and, in fact, it seems reasonable to deny that necessity and, actually, most of the modal notions are definable on the basis of logical consequence, derivability or other

²McGee [105], Chap. 2, follows this line of thought and argues that necessity should be understood as derivability within an explicit system of rules.

notions, which we nowadays attribute to the metalanguage—especially if we think of the modalities in a broader sense as to include propositional attitudes. One might even acknowledge that the modal notions and certain metalinguistic expressions are intimately connected. Yet, there is a distinction to be made between the claim that modal notions should be identified with, or defined by, metalinguistic predicates and the possibility of interpreting modal notions by metalinguistic predicates.³ On the first account modal notions are really just abbreviations of metalinguistic predicates and thus do not appear in our object language. But on the second account we introduce the modal notions into the object language and then try to provide an interpretation of these notions within a metalanguage, i.e., the semantics, where they get interpreted by metalinguistic predicates.

The former option conceives of modalities as predicates of the metalanguage and is commonly viewed as a predicate approach to modality. But one might also label this kind of approach eliminativist, as the modal notions do not occur in the logic proper but only in the metalanguage and, consequently, we cannot formulate modal claims in the language itself.

The latter approach is *prima facie* neutral with respect to the question whether modal notions are to be treated as predicates or operators. That is, one is free to choose either and can, for instance, justify one's choice by referring to our use of the modalities within natural language or, alternatively, which approach proves more convenient. But if the modalities are treated as operators in the object language the approach is not a predicate treatment of modalities, even though the modal notions might get interpreted by predicates.

Carnap in “Die logische Syntax der Sprache” ([26]) and Quine, throughout his writings, clearly advocate the former view, i.e., they conceive of modalities as metalinguistic predicates. For instance, in [127] Quine writes:

Modalities (...) are most clearly construed as properties and relations of statements, and treated in a theory that discourses *about* statements. In such a theory, known nowadays as *metalogue*, variables ‘x’, ‘y’, etc., occur which refer to statements and thus admit not statements but *names of* statements as substituends. Thus the modalities are lifted out of logic proper, and reserved to a discipline that *treats of* the expressions *used in* that logic. (See [127], p. 12. The italics are Quine's.)⁴

However, both acknowledge the possibility of conceiving of modal notions as operators without being confused about use and mention and in his later work Carnap [27] even embraces the latter conception of modalities.⁵

Still, Carnap [26] and Quine think of modal notions as metalinguistic predicates and, accordingly, the sentences of modal operator logic are unintelligible, if taken at face value. To the rescue of modal operator logic Quine argues that we can translate the modal statements involving the modal operator into metalinguistic statements

³Cf. Reinhard [139] discusses this issue to some extent.

⁴Page numbers refer to the reprint in [135].

⁵That is, Carnap [27] introduces modal operator into his object language and provides a semantic interpretation of the operator in terms of a metalinguistic predicate, namely logical truth. However, logical truth is not a primitive notion but definable on the basis of a parametrized truth predicate.

involving a metalinguistic or semantic predicate.⁶ In a nutshell, his proposal is as follows. Let \mathcal{L}_\square be the modal operator language and \mathcal{L}_{Meta} the metalanguage of classical first-order calculus. We set up a translation function I which translates sentences of \mathcal{L}_\square into sentences of \mathcal{L}_{Meta} as follows:

$$I(\phi) := \begin{cases} \phi \text{ is true,} & \text{if } \phi \text{ is atomic} \\ \text{It is not the case that } I(\psi), & \text{if } \phi \doteq \neg\psi \\ I(\psi) \text{ and } I(\chi), & \text{if } \phi \doteq \psi \wedge \chi \\ \psi \text{ is logically true,} & \text{if } \phi \doteq \Box\psi \end{cases}$$

I translates a formula $\Box\chi$ into ‘ χ is logically true’ where logical truth might be replaced by alternative metalinguistic notions depending on the modality under consideration.⁷ Unsurprisingly, this translation highlights most of the caveats Quine and Carnap have with respect to the operator treatment of modality. The first problem is that the modal operator is not truth functional and thus the truth or falsehood of a formula of the form $\Box\chi$ does not solely depend on the truth or falsehood of its subformulas, i.e., χ . As a consequence this provokes the modal operator language to be intensional in the sense that equivalent formulas cannot be substituted *salva veritate* within modal formulas.⁸

The second problem is Quine’s iteration problem. As Quine had repeatedly pointed out, iteration of the modal operator makes it difficult to understand the operator as a metalinguistic predicate.⁹ Remember that via the translation I we wanted to translate sentences of the modal operator language by replacing the modal operator by a metalinguistic predicate. Unfortunately, the function I just defined fails to do so. The translation function is not defined for iterated occurrences of modal operators since if so, it would translate a formula $\Box\Box\chi$ by ‘ $\Box\chi$ is logically true’. Yet, ‘ $\Box\chi$ ’ is not a term of \mathcal{L}_{Meta} . One might try to fix this problem by translating $\Box\Box\chi$ into ‘ χ is logically true’ is logically true’ but this is not an expression of \mathcal{L}_{Meta} either. A metalinguistic predicate is applicable to names of sentences of the object language only but ‘ ϕ is logically true’ is a name of a sentence of the metalanguage. Therefore the translation vindicates only a fragment of the modal operator language and logic.

⁶Sometimes Quine even speaks carelessly of interdefinability (cf. Quine [136], p. 7), which in any ordinary sense of “definability” can’t be right. See again Reinhard [139] for more on this.

⁷In the following discussion ‘ χ ’ is the name of a particular formula although it is arbitrary which formula we choose.

⁸The conception of the distinction between an extensional and intensional language is roughly the one advocated by Quine (cf. Quine [133], p. 151).

⁹Compare, e.g., Quine [130], where he writes:

Grammatically ‘ \Box ’ is an adverb; ‘is analytic’ is a verb. The formal difference becomes immediately apparent in the case of iteration: ‘ \Box ’ can significantly be applied repeatedly (because the result of any application, being a statement, is the kind of expression to which ‘ \Box ’ can still significantly be prefixed) whereas ‘is analytic’ can be applied only once (because the result of the application is a statement, whereas ‘is analytic’ can be appended only to a name).

If one is willing to ascent to a hierarchy of metalanguages and “*extend the notion of validity beyond the semantics of logic to the semantics of semantics*”¹⁰ one can take the outer occurrence of ‘is logically true’ to be of the metalanguage of \mathcal{L}_{Meta} . In this case the modal operator will not get translated into one metalinguistic predicate but into a family of predicates, i.e., the formula $\ulcorner \Box \Box \chi \urcorner$ would get translated by “ χ is logically true $_{\mathcal{L}_{Meta}}$ ’ is logically true $_{\mathcal{L}_{MetaMeta}}$ ’.

The alternative option is to let the predicate ‘is logically true’ apply to sentences of the metalanguage itself. But this undermines the whole idea of a metalanguage and of reducing the modal operator to a metalinguistic predicate. Rather in this case we would have a language which contains names of sentences and a predicate of sentences. This is not a metalinguistic treatment of modality as advocated by Carnap and Quine but rather a predicate or syntactical treatment of modality, which claims that the modal notions are to be construed as predicates within the (object) language.

This shows that metalinguistic treatments of modality along the lines of Carnap and Quine have to either dismiss iterated occurrences of one and the same modal notion as ungrammatical or argue that, despite the initial evidence to the contrary, there is not one but several modal notions involved in the modal statement at stake. This seems problematic from a natural language perspective because we iterate modal notions in natural language and at least *prima facie* it seems that we can even iterate one and the same modal notion. While Carnap and Quine would have concluded that this just highlights the notoriously ambiguous, imprecise and inconsistent character of natural language, this conclusion is less satisfactory from a contemporary perspective. Consequently, if one shares sympathy with predicate treatments of modality one might shy away from the metalinguistic approach to modality, but try to introduce modalities directly as predicates of sentences or propositions into the object language.¹¹ But in Kaplan and Montague [81], and Montague [111], Richard Montague launched an influential attack on predicate treatments of modality.¹² Montague argued that the treatment of a wide class of modal notions as predicates will lead to paradoxical consequences similar to those Tarski had observed with respect to truth. More specifically, variants of Tarski’s undefinability theorem can be proven with respect to the modal notions.

Famously, Tarski [163] has shown that presumably innocent and intuitive assumptions on behalf of the truth predicate will lead to contradiction. He sums up these assumptions in the following renown passage:

(I) We have implicitly assumed that the language in which the antinomy is constructed contains, in addition to its expressions, also the names of these expressions, as well as semantic terms such as the term “true” referring to sentences of this language; we have also assumed that all sentences which determine the adequate usage of this term can be asserted in the language. A language with these properties is called “*semantically closed*”.

(II) We have assumed that in this language the ordinary laws of logic hold.

¹⁰Cf. Quine [130], p. 107. Page numbers refer to the reprint in [135].

¹¹This does not exclude the possibility of understanding the modal notion as logical truth or analyticity but these notions would no longer be conceived as concepts of a metalanguage.

¹²Actually, Myhill [115] had pointed out the paradoxical consequences with respect to the notion of ‘absolute provability’ pretty much parallel to [81].

(III) We have assumed that we can formulate and assert in our language an empirical premise such as statement (2) [roughly, (2) is a sentence which asserts of itself that it is not true, J.S.] which has occurred in our argument. (Tarski [164], p. 348)

Now, the crucial completion of this passage is of course to point out that the adequate usage of the truth predicate is determined by all the sentences of the form:

$$(TB) \quad T\bar{\phi} \leftrightarrow \phi.$$

Here, ϕ is a sentence of the language under consideration and $\bar{\phi}$ a name of ϕ . This latter requirement on the truth predicate is known as Convention T.

Montague observed that it was sufficient for a semantic term α to be characterized by the principles

$$(T) \quad \alpha(\bar{\phi}) \rightarrow \phi$$

$$(Nec) \quad \frac{\phi}{\alpha(\bar{\phi})}$$

for a contradiction to arise.¹³ But (T) and (Nec) are meant to be constitutive principles of a wide range of modal notions, for instance, necessity and are consequently employed in the key systems of modal operator logic. Montague then concluded that syntactical treatments of modality—which was his label for predicate approaches to modality—cannot embrace the fundamental modal principles, which he took to be a strong argument in favor of the operator approach.

We will soon take a closer look at Montague’s theorem and the conclusion one should draw from it but what is puzzling from the historical perspective is that there seems to be no defense of a “syntactical” approach to this date in the literature with exception of the work by Carnap and Quine. And to be sure, Montague names Carnap and Quine as his main targets¹⁴ even though, as we have argued, Carnap and Quine advocate a metalinguistic approach to modality. Thus Montague’s theorem does not apply to their proposal as their language is not semantically closed.¹⁵

Of course, there had been further proposals of so-called quotational approaches to indirect discourse which construct modal locutions as ‘Peter believes that S ’ as a relation between an agent, Peter, and the sentence S or an entity which shares the syntactical structure of S . In general, proponents of quotational approaches to indirect discourse analyze indirect discourse by appeal to direct discourse, i.e., the

¹³From a natural language point of view the difference between the right-to-left direction of (TB) and (Nec) relies in hypothetical reasoning. If we assume ϕ , then by (TB) we are licensed to infer $T\bar{\phi}$. In the case of (Nec) we can infer $T\bar{\phi}$ only if ϕ is true.

¹⁴See Montague [111], footnote 1.

¹⁵See Niebergall [116] for remarks along these lines. To be fair, Quine sometimes seems to suggest that in order to make sense of iterated uses of the modal operator we should apply the semantically predicate ‘Nec’ to sentences in which it occurs (e.g., in Quine [131]). But we think a charitable reading of Quine reconstructs him as advocating the hierarchy solution. However, Reinhardt [139] thinks differently.

sentence ‘Peter believes that *S*’ would be analyzed in a way similar to the sentence ‘Peter believes ‘*S*’’. That is to say, by the modal predicate, ‘Peter believes’, which is applied to the name of a sentence, ‘*S*’. Proposals of this kind had been advocated by Carnap [27], Putnam [126] and Scheffler [146] before Montague’s results. But it is unclear whether they intended their proposals to be metalinguistic or not and whether Montague meant to attack these proposals.¹⁶

Be that as it may, due to the work of Montague it became clear that philosophers willing to embrace a quotational approach to indirect discourse and logicians hoping to treat modalities as predicates will have to deal with paradoxes—paradoxes similar to those the truth theorist has to face—as long as they do not opt for a metalinguistic treatment. Moreover, in subsequent work Germano [48], Friedman and Sheard [47] and Thomason [168], amongst others, have shown that further combinations of modal principles might lead to inconsistency. Importantly, Thomason showed that the paradoxes affect formal treatments of propositional attitudes like belief. Using his and Montague’s result Thomason [166, 168, 169] argued that quotational or direct discourse approaches to propositional attitudes are affected by the paradoxes.¹⁷ Egré [38] provides a good overview of the different inconsistency results but see also our next chapter.

Surprisingly, only after Montague’s results syntactical or predicate treatments of modality, as opposed to metalinguistic treatments, received serious attention as can be witnessed, for example, by the work of Germano [48], Anderson [5], Perlis [121, 122] and, Asher and Kamp [10, 11]. However, within mainstream philosophy it seems that Montague’s theorem was pretty much taken to show that modalities ought to be treated as operators and that the syntactic approach to modality was a failure. Thus, e.g., Kripke writes¹⁸:

Montague and Kaplan, and Montague, following Gödel, have proven sharp results showing that if propositional attitudes or modalities are treated as properties of sentences paradox will result unless special precautions are taken. (See [92], p. 414).

Considering that Montague’s theorem is a variant of Tarski’s undefinability theorem this reception of Montague comes surprising, as it diverges to great extent to the reaction Tarski and, following Tarski, the scientific community, showed with respect to the undefinability of truth. Famously, Tarski himself did not question whether truth was aptly formalized by a predicate but concluded that truth was an expression of the metalanguage. Accordingly, he gave up the semantical closedness of the language. Sentences in which the truth predicate is applied to sentences of the metalanguage are

¹⁶If Montague did mean to attack these proposals, it is bizarre that he did not mention the work of Church [29] who already attacked these proposals or more generally “sententialism” on more philosophical grounds. Thomason [166] takes Montague’s result to be, in some sense, a sequel to Church’s criticism of sententialism.

¹⁷Specifically, Thomason means to include representationalist theories of propositional attitudes.

¹⁸An even more pronounced statement can be found in Slater [153] who writes:

Since Montague, we surely now know that syntactic treatments of modality must be replaced by operator formulations. (See [153]. p. 453).

either banned from the language by syntactic means, i.e., the notion of a well-formed formula is thus defined that the truth predicate cannot be concatenated with names of sentences of the metalanguage or, alternatively, sentences of this kind are globally declared as false.¹⁹ Carnap's [26] and Quine's approach to modality can therefore be seen as essentially Tarskian in spirit and Carnap [26] even argues that the concept of analyticity, which he takes to be a possible *definiens* of necessity, cannot be defined within the language itself for sake of paradox.

The technical motivation for allowing the truth predicate and the modalities to be predicates of the metalanguage only is, of course, that one can thereby block the construction of self-referential, liar like sentences. For instance, using the Gödelian techniques, if truth were a predicate of the object language and applicable to sentences containing the truth predicate, then we could construct a sentence saying of itself that it is not true. That is, we can find a sentence ' λ ' such that ' $\neg T\bar{\lambda} \leftrightarrow \lambda$ ' is provable in some basic arithmetical theory such as Q where ' $\bar{\lambda}$ ' would be ' $\overline{gn(\lambda)}$ ', i.e., ' $\lceil \lambda \rceil$ ' according to our terminology. But then, assuming that the truth predicate respects the principle (TB), the existence of such a sentence λ leads to inconsistency relative to the arithmetical theory.

In his seminal "Outline of a theory of truth" Kripke [91] challenged the metalinguistic account of truth. He criticized Tarski's proposal for being unprincipled, as a diagnosis of the paradoxes reveals that the paradoxality does not depend on structural, i.e., syntactic, features of certain sentences. To stick to Kripke's own words the metalinguistic or hierarchical approach to truth "seems unfaithful to the facts",²⁰ i.e., natural language.

In a nutshell Kripke's argument is that there are iterated uses of truth in natural language and that it does not seem correct to assign to sentences a level in the hierarchy of metalanguages at the outset depending solely on syntactic or structural features of the sentence. Since it is hard to out-rule Kripke in brevity and clarity we shall cite a longer passage of his to establish the point:

If someone makes such an utterance as (1) [Most, i.e., a majority, of Nixon's statements about Watergate are false. (J.S.)] he does *not* attach a subscript, explicit or implicit, to his utterance of '*false*', which determines the "level of language" in which he speaks. An implicit subscript would cause no trouble, if we were sure of the "level" of Nixon's utterance; we could then cover them all in the utterance of (1) or even of the stronger

(4) All of Nixon's utterances about Watergate are false.

simply by choosing a subscript higher than the levels of any involved in Nixon's Watergate related utterances. Ordinarily, however, a speaker *has no way of knowing the "levels" of Nixon's relevant utterances.* (...) The higher the "levels" of Nixons utterances happen to be, the higher the "level" of (4). This means that in some way a statement should be allowed to seek its own "level", high enough to say what it intends to say. It should not have an intrinsic level fixed in advance as in the Tarski hierarchy. (See Kripke [91], pp. 695/696.)

¹⁹The former approach is Tarski's original proposal, the latter can, for instance, be found in Halbach [60, 66].

²⁰Cf. Kripke [91], p. 695.

After dismissing the hierarchical, i.e., metalinguistic approach to truth Kripke proposes a weakening of the adequacy condition of truth put forward by Tarski. However, Kripke did by no means consider treating truth as an operator in order to avoid the paradoxes.²¹ Now, some authors have defended an operator view of truth, but those authors (or at least the reasonable amongst them) have been motivated by philosophical and linguistic consideration and don't conceive of their proposal as a reaction to the paradoxes.²²

It thus seems that Skyrms is quite right in stating:

No one, I hope, has taken Tarski's theorem as a proof that truth is not a predicate of sentences. Tarski's theorem is not a theorem of the redundancy theory of truth. (Skyrms [152], p. 368)

On the face of it there seems to be no reason why the reaction toward Montague's theorem should be any different, if no strong case for the disanalogy between the case of truth and the modalities is made. And most authors advocating predicate approaches to modality embrace Kripke's argumentation to a large extent, that is, they take self-reference to be an important characteristic of natural language—something which should be reproduced within a formal setting rather than artificially banned from the formal language—and thus accept weaker adequacy criteria for the modal notions than the standard laws of modal operator logic.

This seems to be the more positive understanding of the moral Montague draws from his result, as he states:

Thus if necessity is to be treated syntactically, that is, as a predicate of sentences, as Carnap and Quine have urged, then virtually all of modal logic, (...), must be sacrificed. (See Montague [111], p. 294. Page numbers refer to the reprint in Montague [112].)

Montague takes this to be an argument in favor of the operator approach and opts for an operator treatment of modalities and, consequently, does not take self-referential modal sentences to be an important feature of a formal account of the modalities. Apparently, many philosophers have followed him on this behalf and so the question arises whether something can be said to the rescue of predicate approaches to modality, if the unrestricted characteristic principles employed in modal operator logic such as (*T*) and (*Nec*) are conceived as indispensable for any adequate formalization of the modalities.

2.2 Montague's Theorem and Modal Logic

As we have pointed out, common wisdom suggests that modalities need to be treated as operators in order to avoid paradoxical consequences once key modal principles are

²¹ And this would of course violate the spirit of his approach. Kripke's main idea is that we should accept self-reference as an essential feature of our language and formalizing truth by an operator seems to eradicate all self-reference with respect to truth from the language.

²² Amongst those authors who champion an operator approach to modality are Prior [125] and more recently Mulligan [114]. Ramsey's redundancy theory of truth and Grover's et al. [54] prosentential theory of truth are often understood as advocating an operator approach.

postulated.²³ But before we follow common wisdom and Montague in this conclusion a closer look at Montague's results can be of no harm. The results established in Kaplan and Montague [81] and Montague [111] rely on the capacity of the theory under consideration to prove the Diagonal lemma and to encode the syntax of the language via Gödel numbering. Within the arithmetical setting Robinson Arithmetic, i.e., Q suffices for this purpose. Moreover, via Gödel numbering we possess a name for every expression of the language, namely, the numeral of the Gödel number of this expression. As a consequence, every theory in which Q can be relatively interpreted will have the means necessary to encode syntax and to prove the Diagonal lemma. Now, Montague shows that any theory Σ is inconsistent, in which Q can be relatively interpreted and where a formula $\alpha(x)$ is governed by the modal principle (T) and the rule (Nec) .

However, for expository ease we state Montague's theorem in a slightly less general way and omit the complication of relative interpretability. Instead we require a theory Σ to be an extension of Q in the language of arithmetic or an extension thereof. But we start by presenting Tarski's undefinability theorem for this will allow us to view Montague's theorem as a straightforward strengthening of Tarski's:

Theorem 2.1 (Tarski/Gödel) *Let Σ be a theory extending Q in $\mathcal{L}_{PA\&}$ and ' α ' a (possibly complex) one-place predicate. If for every sentence $\phi \in \mathcal{L}_{PA\&}$*

$$(i) \quad \Sigma \vdash \alpha(\ulcorner \phi \urcorner) \leftrightarrow \phi$$

then Σ is inconsistent.

Proof As a consequence of the Diagonal lemma there is a sentence λ such that

$$\Sigma \vdash \lambda \leftrightarrow \neg \alpha(\ulcorner \lambda \urcorner)$$

but since by (i) $\Sigma \vdash \alpha(\ulcorner \lambda \urcorner) \leftrightarrow \lambda$ the contradiction is immediate. \square

Montague's theorem can now be presented as a strengthening of the above theorem. That is, one direction of (TB) is replaced by a rule of inference but the theory remains inconsistent. The inconsistency can then be derived using the same sentence λ as in the above proof

Theorem 2.2 (Montague) *Let Σ be a theory extending Q in $\mathcal{L}_{PA\&}$ and ' α ' a (possibly complex) one-place predicate. If for every sentence $\phi \in \mathcal{L}_{PA\&}$*

$$(i) \quad \Sigma \vdash \alpha(\ulcorner \phi \urcorner) \rightarrow \phi$$

$$(ii) \quad \Sigma \vdash \phi \Rightarrow \Sigma \vdash \alpha(\ulcorner \phi \urcorner)$$

then Σ is inconsistent.

²³As we have mentioned in the Acknowledgements the material of this section has already been published as Stern [159].

Proof Again the proof makes essential use of the Diagonal lemma:

- | | | |
|----|---|----------------|
| 1. | $\Sigma \vdash \neg\alpha(\ulcorner\lambda\urcorner) \leftrightarrow \lambda$ | Diagonal lemma |
| 2. | $\Sigma \vdash \alpha(\ulcorner\lambda\urcorner) \rightarrow \lambda$ | (i) |
| 3. | $\Sigma \vdash \neg\alpha(\ulcorner\lambda\urcorner)$ | 1, 2 |
| 4. | $\Sigma \vdash \lambda$ | 1, 3 |
| 5. | $\Sigma \vdash \alpha(\ulcorner\lambda\urcorner)$ | 4, (ii) |

□

These inconsistency results obviously rely on the possibility to find such a sentence λ and thus on the possibility to diagonalize the predicate α . If one were to block the possibility of diagonalizing α , the paradox could no longer be derived along the outlines of Theorems 2.1 and 2.2.

Basically, this is the strategy of Tarski's and other metalinguistic proposals which block iterated applications of the predicate α . In the above derivation λ is the sentence

$$\neg\alpha(sub^\bullet(\ulcorner\neg\alpha(sub^\bullet(v_0, v_0))\urcorner), \ulcorner\neg\alpha(sub^\bullet(v_0, v_0))\urcorner))$$

and therefore λ cannot be constructed, if iterations of α are prohibited by the formation rule of the language. However, Skyrms [152] develops an approach in this spirit, which does allow for iteration of the modal predicate but while nonetheless the diagonalization of the modal predicate is blocked. Still, Skyrms alters the formation rules of the language and imposes syntactic restrictions on the notion of a formula. Skyrms' language is constructed starting from the arithmetical language by a recursion over the natural numbers. At each step names of the sentences constructed at the previous step are added to which the modal predicate can be applied to form new formulas. These names are not ordinary terms of the language as no predicate other than the modal predicate can be applied to form a new sentence. Also, the modal predicate is not an ordinary predicate, as it is applicable to names of sentences only and not to other terms of the language. Especially, the predicate is not applicable to variables and thus it is not possible to quantify into the argument position of the modal predicate. At each step, the set of formulas is closed under the boolean operations, that is, besides the formulas constructed at previous steps, the set of formulas contains the newly constructed modal formulas and all the resulting boolean combinations of the formulas at this step. Ultimately, Skyrms takes the modal language to be the union of all the languages constructed in the recursion process and shows that in this setting we may have a theory extending Q which is closed under (*Nec*) and in which (*T*) is a theorem. That is, we can postulate the classical principles of modal operator logic in the predicate setting without falling prey to paradox. Skyrms

even shows that his account of modality is adequate with respect to a certain modal operator logic: the theorems of this modal operator logic are, modulo translation, exactly the valid formulas of his account.

As in the case of Tarski's proposal the sentence λ cannot be constructed because α applied to a term $sub^\bullet(x, y)$ is not a formula of the language. This restriction obviously comes at a cost. Namely, if we conceive of the names of formulas as the numerals in the standard Gödelian fashion,²⁴ then the resulting logic will not be closed under substitution, be it substitution *salva veritate* or even *salva congruitate*: for instance, $\alpha(\lceil S(\bar{0}) = \bar{1} \rceil)$ is a formula but $\alpha(sub^\bullet(\lceil S(v_0) = \bar{1} \rceil, \bar{0}))$ is not, even though clearly $\lceil S(\bar{0}) = \bar{1} \rceil = sub^\bullet(\lceil S(v_0) = \bar{1} \rceil, \bar{0})$.

Skyrms introduces new names for the sentences which are distinct from the numerals and thereby blocks the unwelcome consequence we have just sketched. But these names are highly suspect, as they do not behave like ordinary terms of the language. They only occur in the argument position of the modal predicate and, in fact, it is not obvious why we should not read the modal predicate applied to these names of sentences as a modal operator applied to the sentences themselves. We shall come back to this issue, but the upshot of this discussion is that it is unclear whether we should really take Skyrms' modal predicate to be a predicate and not an operator in disguise.

Be this as it may, the merit of Skyrms' proposal is to point out that it is not as straightforward as Montague would have it to conclude from his theorems that all of modal logic must be sacrificed once modalities are treated as predicates (of sentences). Rather this depends on further assumptions, for instance, that the modal predicate can be diagonalized.

Blocking the diagonalization of the modal predicate amounts to imposing syntactic restrictions on the formation rules of the language under consideration and, consequently, this strategy always is *ad hoc* in flavor. For how could such a restriction be justified on philosophical grounds?

But Skyrms' approach also points to another direction. Namely, *prima facie* there does not seem to be any philosophical justification or requirement for the names of the sentences to be the numerals of the Gödel codes of the sentences—which we shall call the arithmetical names from now on—as assumed in Montague's theorem. And this fact has been exploited by Niemi [117], Gupta [57] and Schweizer [148]. Gupta, for instance, shows that once we postulate a distinct class of quotation names which are used to state the principle (TB), we can construct a classical model for truth, that is, we can have, contra Tarski, a semantically closed language in which the truth predicate respects the principle (TB). As we shall see, Gupta's strategy generalizes to the modal case and can even be used to provide a predicate account of multimodal logics.

²⁴This is essentially the proposal made by Belnap and Gupta [14] when they present their language \mathcal{PL}^- (cf. pp. 241–243) and indirectly also affects the proposal of Des Rivières and Levesque [37].

2.2.1 A Classical Model for Truth

Gupta's construction relies on an observation which was probably noted explicitly for the first time by Niemi [117] who provided a predicate account of modalities embracing, amongst other principles, the modal principles (T) and (Nec), which according to Montague's theorem lead to an outright contradiction in the predicate setting. But Niemi showed the consistency of his modal theory. This was achieved by stipulating a distinct class of names for the sentences of the language and by using these names to state his modal axioms. Contrary to the arithmetical names which serve as the names of the sentences in Montague's formulation of the modal principles these newly introduced names of sentences do not necessarily provide the resources to represent syntax, i.e., concepts of syntax such as the substitution function. But these resources are of crucial importance in the derivation of the modal antinomies or the undefinability of truth.

This reveals a further often unnoticed assumption the derivation of Tarski's and Montague's theorem relies on which is nicely summed up by Schweizer:

There are really two independent assumptions built into the above phenomenon of modal self-reference:

- (a) the possession of a class of terms structurally rich enough to do arithmetic and to sustain the Diagonal lemma, and
- (b) the use of these terms as the privileged names of syntactical objects in defining the modal logic. (Schweizer [148], p. 7)

Let us observe how the derivation of Tarski's theorem can be blocked if assumption (b) is dropped. As language we may suppose the language \mathcal{L}_{PA} supplemented by a class of quotation names for the sentences of the language and a truth predicate. These quotation names are conveyed by formulas flanked by squiggly quotes, i.e., $\mathcal{I}\phi\mathcal{I}$ is the name of the sentence ϕ .²⁵ Furthermore we shall assume a theory Σ extending Q that proves the following version of (TB)

$$(TB) \quad T\mathcal{I}\phi\mathcal{I} \leftrightarrow \phi$$

Our theory Σ still has the means to prove the Diagonal lemma and thus we can find a sentence λ for which Σ proves

$$(L) \quad \neg T[\lambda] \leftrightarrow \lambda$$

But this by no means licenses us to derive a contradiction considering the way (TB) was stated. We can only derive

²⁵These quotes, i.e., ' \mathcal{I} ' and ' \mathcal{I} ' are to be distinguished from Quine corners, ' \ulcorner ', ' \urcorner ' and also our Gödel corners ' \lceil ', ' \rceil '. The former are part of the object language, whereas the Quine corners are not and the Gödel corner are used to represent the numeral of the Gödel number of the sentence flanked by our Gödel corners and thus they are no symbols of the language in their own right.

$$\neg T[\lambda] \leftrightarrow T\mathfrak{J}\lambda\mathfrak{J}$$

In order to derive a contradiction we would need to substitute $[\lambda]$ for $\mathfrak{J}\lambda\mathfrak{J}$ above or conversely. But such a substitution will only be licensed if

$$(GQ) \quad [\lambda] = \mathfrak{J}\lambda\mathfrak{J}$$

and this identity statement should by no means be taken as a given. Rather one motivation for introducing a class of quotation names might be to block the common identification of sentences with their codes, i.e., their Gödel numbers. In this case (GQ) should not be expected to be provable in a theory Σ nor should one expect (GQ) to be a true statement since—as mentioned—the two terms will refer to two different objects. Alternatively, one might suspect that a statement parallel to (L) but where the occurrence of the code of λ is substituted by the quotation name of λ , i.e.,

$$(QL) \quad \neg T\mathfrak{J}\lambda\mathfrak{J} \leftrightarrow \lambda$$

can be proven within Σ . Yet, as we have pointed out, one should not expect (QL) to be trivially provable in Σ since contrary to the case of the arithmetical terms we do not know whether our quotation names “are structurally rich enough to do arithmetics and to sustain the Diagonal lemma”. Accordingly, one might introduce quotation names in order to work within a setting in which names for the sentences of the language are available without being prone to the effects of diagonalization. In fact this is precisely the framework of the accounts of Niemi, Gupta and Schweizer.

In contrast to Niemi, Gupta does not construct a particular theory but shows that independently of the base language and theory assumed we can construct a classical model for truth in which the principle (TB) holds unrestrictedly. The construction shows that Tarski’s Convention T can be satisfied in a semantically closed language, that is, we can construct a model in which (TB) holds.

Gupta’s basic strategy is to start with a model in which the truth predicate is assigned an arbitrary extension and then revise this extension in a sequence of steps, i.e., the extension of the truth predicate at step $\alpha + 1$ is the set of sentences true in the model at step α . At limit ordinals all sentences, which have remained stably in the extension of the truth predicate from an ordinal $\beta < \alpha$ onward, are gathered to build the extension of T at the limit ordinal. We know that this process will not in general lead to a classical model of the language in which (TB) holds unrestrictedly as long as no further assumption on behalf of the interpretation of the quotation names is made. As we have seen it needs to be guaranteed that

- the denotatum of the quotation name of a sentence is not the denotatum of the arithmetical name of the sentence—its Gödel number for instance—and
- no function symbol or predicate is interpreted as a function or relation on the denotata of the quotation names in a way that allows us to interpret a sentence of the language as the liar sentence or some related paradoxical sentence.

Gupta shows that any initial model in which these two conditions are met can be extended to a model in which (TB) holds unrestrictedly. That is, Gupta shows that in this case (all) the sequence(s) of revisions of the interpretation of the truth predicate converge(s) to a unique fixed-point at the first limit ordinal ω .

While the first condition can be imposed on the interpretation function in a straightforward manner, it is unclear what we need to require exactly of the interpretation in order for the latter condition to be satisfied. It is clear, however, that if we do not allow the interpretation of predicates and function symbols other than the truth predicate to discriminate between sentences, then it will be satisfied. This condition is clearly too strong and can be liberalized in different ways. For instance, the interpretation of certain predicates can discriminate between sentences and we can thus allow a predicate saying of a sentence that it is the negation of another sentence and, similarly, predicates for the remaining truth functional connectives. To our knowledge, it remains an open question how to spell out the second condition in its most liberal fashion, but to guarantee that the revision sequences still converge to a unique fixed point.

We shall sketch Gupta's construction assuming the most restrictive condition on the interpretation function of models²⁶ but will pause to indicate how the construction can be altered in order to accommodate sentential predicates for the truth functional connectives. We start by defining the language \mathcal{L}_{QT} which has the peculiarity of possessing besides the usual terms and predicates a class of quotation names and a truth predicate. The vocabulary consists of the symbols of an arbitrary first-order language augmented by the quotation symbols ' \ulcorner ' and ' \urcorner ' and the truth predicate ' T '. As one would suspect, in the presence of quotation names the expressions 'formula' and 'term' are defined by a simultaneous induction.

Definition 2.3 (*Term, Formula and Quotation degree*) The expressions 'term', 'formula' and 'quotation degree' of \mathcal{L}_{QT} are defined simultaneously. The quotation degree is a function $qd : Term_{\mathcal{L}_{QT}} \cup Frml_{\mathcal{L}_{QT}} \rightarrow \omega$ assessing the depth of embeddings of quotations in a formula

1. If t is a variable or an individual constant, which is not a quotation name, then $t \in Term_{\mathcal{L}_{QT}}$ with $qd(t) = 0$;
2. if $t_1, \dots, t_n \in Term_{\mathcal{L}_{QT}}$ and f^n is a function symbol, then $f^n(t_1 \dots t_n) \in Term_{\mathcal{L}_{QT}}$ with $qd(f^n(t_1 \dots t_n)) = \max(qd(t_1), \dots, qd(t_n))$;
3. if t_1, \dots, t_n are terms of \mathcal{L}_{QT} and P^n an n -ary predicate constant, then $P^n t_1, \dots, t_n \in Frml_{\mathcal{L}_{QT}}$ with $qd(P^n t_1, \dots, t_n) = \max(qd(t_1), \dots, qd(t_n))$;
4. if $\phi \in Frml_{\mathcal{L}_{QT}}$, then $\neg\phi$ and $\forall x\phi$ are formulas of \mathcal{L}_{QT} with $qd(\neg\phi) = qd(\forall x\phi) = qd(\phi)$;
5. if ϕ and ψ are formulas of \mathcal{L}_{QT} , then $\phi \wedge \psi \in Frml_{\mathcal{L}_{QT}}$ with $qd(\phi \wedge \psi) = \max(qd(\phi), qd(\psi))$;
6. if $\phi \in Sent_{\mathcal{L}_{QT}}$, then $\ulcorner\phi\urcorner \in Term_{\mathcal{L}_{QT}}$ with $qd(\ulcorner\phi\urcorner) = qd(\phi) + 1$.

As mentioned, only certain models can be extended to a classical model satisfying (TB). We call models that qualify in this respect proper premodels.

²⁶Our presentation closely follows Gupta [57], pp. 9–15.

Definition 2.4 (*Proper premodel*) A proper premodel of \mathcal{L}_{QT} is a tuple (D, I) with $Sent_{\mathcal{L}_{QT}} \subset D$ and an interpretation function I on the whole vocabulary of \mathcal{L}_{QT} except the truth predicate such that I has the following properties:

1. $I(\ulcorner \phi \urcorner) = \phi$ for every sentence ϕ ;
2. if a term t is not a quotation name, then $I(t) \notin Sent_{\mathcal{L}_{QT}}$;
3. if P is an n -place predicate and d_i is a sentence of \mathcal{L}_{QT} for $1 \leq i \leq n$, then $(d_1, \dots, d_i, \dots, d_n) \in I(P)$ iff for all $d'_i \in Sent_{\mathcal{L}_{QT}}$, $(d_1, \dots, d'_i, \dots, d_n) \in I(P)$.
4. No sentence appears in the range of $I(f)$, f being an n -place function symbol. If $d_i, d'_i \in Sent_{\mathcal{L}_{QT}}$, $1 \leq i \leq n$, then $I(f)(d_1, \dots, d_i, \dots, d_n) = I(f)(d_1, \dots, d'_i, \dots, d_n)$.

We denote the class of all proper premodels by \mathfrak{M} .

Definition 2.4 reflects the informal condition set out before. Item 1 and 2 guarantee that the quotation name of a sentence and its arithmetical name will not refer to the same object, i.e., we distinguish between the sentence and its code. Items 3 and 4 on the other hand guarantee that the second condition is met. The interpretation of predicates different than ‘ T ’ does not discriminate between different sentences. Either no sentence appears as a relatum in the interpretation of a predicate or all of them do. The same holds for the interpretation of a function symbol but, additionally, no sentence is allowed in the range of a function interpreting a function symbol. The definition guarantees that no sentence is viewed as the liar sentence or some other paradoxical sentence and, intuitively, this is the reason why we can transform the model into a model in which (TB) is satisfied.

By Definition 2.4 a premodel of the language \mathcal{L}_{QT} is a tuple (D, I) , which, as of yet, does not assign an interpretation to the truth predicate. To obtain a full fledged model of the language \mathcal{L}_{QT} we need to provide some interpretation of the truth predicate. Such a full fledged model for the language \mathcal{L}_{QT} is denoted by (M, S) where M is a proper premodel and S , the interpretation of the truth predicate, is an arbitrary subset of the domain.²⁷ Now, the idea is to start with some arbitrary subset of the domain and then through a series of revisions to obtain an interpretation of the truth predicate with some desirable properties. To this end we shall define a classical jump or revision operation Ξ_M on the domain of a proper premodel M .

Definition 2.5 (*Jump relative to a proper premodel*) Let M be a proper premodel with domain D . Then $\Xi_M : P(D) \rightarrow P(D)$ is a jump operation relative to M iff for all $S \subseteq D$

$$\Xi_M(S) := \{\phi \in Sent_{\mathcal{L}_{QT}} : (M, S) \models \phi\}$$

We may iterate applications of Ξ_M to a given set $S \subseteq \omega$ and thereby obtain a series of revisions of the interpretations of the truth predicate. We define by transfinite recursion:

²⁷In general, if M is a model for a language \mathcal{L} without the truth predicate, then we take (M, S) to be a model of the language \mathcal{L}_T , which we obtain from \mathcal{L} by adding a truth predicate.

Definition 2.6 Let M be a proper premodel, Ξ_M a jump relative to M , and $S \subseteq D$. Then we define iterative applications of Ξ_M to S for ordinals α as follows

$$\Xi_M^\alpha(S) := \begin{cases} S, & \text{if } \alpha = 0 \\ \Xi_M(\Xi_M^\delta(S)), & \text{if } \alpha = \delta + 1 \\ \{\phi \in \text{Sent}_{\mathcal{L}_{QT}} : \exists \kappa(\phi \in \bigcap_{\kappa \leq \beta < \alpha} \Xi_M^\beta(S))\}, & \text{if } \alpha \text{ is a limit ordinal} \end{cases}$$

As mentioned, it can be shown that this process reaches a fixed-point at the first limit ordinal ω . The following lemma is crucial for establishing this fact.

Lemma 2.7 *Let M be a proper premodel with domain D . Then for all $S, S' \subseteq D$, all natural numbers n , and all ordinals α , if $\alpha > n + 1$, then for all sentences ϕ with $dq(\phi) \leq n$:*

$$\phi \in \Xi_M^{n+2}(S) \Leftrightarrow \phi \in \Xi_M^\alpha(S')$$

The lemma establishes that relative to a proper premodel the truth of a sentence of quotation degree n is settled latest at stage $n + 2$. From this point on a sentence is either stably in the extension of the truth predicate or it is stably not in the extension independently of the choice of the initial interpretation of the truth predicate.

Proof Sketch Gupta [57] (pp. 9–15) gives a detailed proof of the lemma. We confine ourselves to giving the crucial ideas of the proof. The proof goes by induction over n and uses a side induction over α . Furthermore, three cases can be distinguished with respect to α . Either (i) α is zero, or (ii) α is a limit ordinal, or (iii) α is a successor ordinal $\delta + 1$. The first two cases are trivial—(ii) due to Definition 2.6 and the induction hypothesis (induction on α).

For (iii) we observe that by Definition 2.6 $\Xi_M^{n+2}(S)$ and $\Xi_M^{\delta+1}(S')$ coincide on the sentences ϕ of quotation degree $\leq n$ iff

$$(M, \Xi_M^{n+1}(S)) \models \phi \Leftrightarrow (M, \Xi_M^\delta(S')) \models \phi.$$

Again by induction hypothesis (induction on n) we know that the two models agree on the sentences of quotation degree $< n$. On the other hand the set $\Xi_M^{n+1}(S)$ partitions the sentences of degree $\geq n$ in two distinct denumerable sets. This is due to the fact that $(M, \Xi_M^{n+1}(S))$ is a classical model and thus there are denumerable many tautologies and denumerable many contradictions of degree $\geq n$. The same holds for $(M, \Xi_M^\delta(S'))$. This allows us to define a mapping σ from the domain onto itself which respects the interpretation of the truth predicate, i.e.,

$$\phi \in \Xi_M^{n+1}(S) \Leftrightarrow \sigma(\phi) \in \Xi_M^\delta(S')$$

where σ is the identity function on all the members of the domain, which are not sentences of degree $\geq n$. In virtue of Definition 2.4 and the fact that any quotation name of quotation degree $\leq n$ has a sentence of degree $< n$ as its denotatum this

guarantees that we can establish for any formula $\phi(x_1, \dots, x_n)$ of quotation degree $\leq n$ and $d_1, \dots, d_n \in D$

$$(M, \Xi_M^{n+1}(S)) \models \phi[d_1, \dots, d_n] \Leftrightarrow (M, \Xi_M^\delta(S')) \models \phi[\sigma(d_1) \dots, \sigma(d_n)].$$

But then as a corollary $(M, \Xi_M^{n+1}(S))$ and $(M, \Xi_M^\delta(S'))$ agree on the sentences of quotation degree $\leq n$ and, consequently, so do $\Xi_M^{n+2}(S)$ and $\Xi_M^\alpha(S')$.²⁸ \square

Remark 2.8 (Truth functional connectives) A closer examination of the proof shows that Definition 2.4 can be liberalized for it to allow for certain predicates to discriminate between sentences of the language. More precisely, we can allow for predicates for the truth functional connectives. That is, our language can contain predicates saying of a sentence that it is the negation of another sentence or saying of a sentence that it is the result of some truth functional operation on some other sentences.

Instead of clause (3) of Definition 2.4 we set

(3⁺) if P is an n -place predicate, then either

- (a) $P = \text{Neg}$ (with arity 2) and $(d_1, d_2) \in I(\text{Neg})$ iff $d_1 = \neg d_2$ with $d_1, d_2 \in \text{Sent}_{\mathcal{L}_{QT}}$.
- (b) $P = \text{Jun}$ (with arity 3) and $(d_1, d_2, d_3) \in I(\text{Jun})$ iff $d_1 = d_2 \mathbf{J} d_3$ where \mathbf{J} is the two place truth functional connective intended to be represented by Jun and $d_1, d_2, d_3 \in \text{Sent}_{\mathcal{L}_{QT}}$.
- (c) P is some other predicate with arity n , then $d_i \in \text{Sent}_{\mathcal{L}_{QT}}$ for $1 \leq i \leq n$ $(d_1, \dots, d_i, \dots, d_n) \in I(P)$ iff for all $d'_i \in \text{Sent}_{\mathcal{L}_{QT}}$, $(d_1, \dots, d'_i, \dots, d_n) \in I(P)$.

If we assume clause (3⁺) instead of (3), we need to construct the bijection σ more carefully. It has to be guaranteed that the bijection respects the interpretation of T but, in addition, it also need to respect the interpretation of Neg and Jun . This will be the case, if σ commutes with the truth functional connectives. Thus σ needs to respect the following conditions:

$$\begin{aligned} \sigma(\neg\phi) &= \neg\sigma(\phi) \\ \sigma(\phi \mathbf{J} \psi) &= \sigma(\phi) \mathbf{J} \sigma(\psi) \end{aligned}$$

for all $\phi, \psi \in \text{Sent}_{\mathcal{L}_{QT}}$. Fortunately, σ can be constructed in a way for these conditions to be met. Since σ is the identity function on the sentences of degree $< n$ σ obviously respects the above conditions on these sentences. Furthermore, we can enumerate all atomic sentences of degree $\geq n$ as well as those sentences of degree $\geq n$ whose main logical symbol is a quantifier. Call the set of all these sentences D^- . Note, that amongst these sentences denumerably many fall under $\Xi_M^{n+1}(S)$ (for some set $S \subseteq D$) and denumerably many sentences do not and similarly for $\Xi_M^\delta(S')$. That is, we can

²⁸ σ needs to respect the interpretation T as, e.g., a sentence of the form $\forall xTx$ is of quotation degree 0.

define a mapping π from D^- onto itself such that $d \in \Xi_M^{n+1}(S) \Leftrightarrow \pi(d) \in \Xi_M^\delta(S')$ and define σ as follows:

$$\sigma(\phi) := \begin{cases} \phi, & \text{if } dq(\phi) < n \\ \pi(\phi), & \text{if } \phi \in D^- \\ \neg\sigma(\psi), & \text{if } \phi \doteq (\neg\psi) \\ \sigma(\psi)\mathbf{J}\sigma(\chi), & \text{if } \phi \doteq (\psi\mathbf{J}\chi) \end{cases}$$

Moreover, σ is the identity function on the $D - \text{Sent}_{\mathcal{L}_{QT}}$. It is now straightforward to check that σ is a bijection on D , which preserves the interpretation of all the predicates—including those of T , Neg and Jun . Given such a bijection σ Lemma 2.7 can be proven as before.

Remark 2.9 One might hope that something like the above can be done with respect to a predicate $true(x, y)$ saying that x is the T -necessitation of y , as intuitively this does not provoke an infinite descending chain or regress but (as the formulas of our language are of finite length) only finite ones.

However, the hope to prove an analogue of Lemma 2.7 without altering, at least, the definition of the quotation degree of a formula are dim. To see why, let us suppose we are constructing a bijection σ from D onto itself respecting the interpretation $true(x, y)$. In order to guarantee this the following would need to hold

$$(\dagger) \quad \sigma(T\ulcorner\phi\urcorner) = T\ulcorner\sigma(\phi)\urcorner$$

for every sentence ϕ of the language. Let us recall that σ is also meant to respect the interpretation of the truth predicate. Unfortunately these two demands on σ are difficult to fulfill simultaneously. Suppose $qd(\phi) = m > n$ and suppose furthermore that $\phi \in \Xi_M^{n+1}(S)$ iff $\sigma(\phi) \in \Xi_M^\delta(S')$. Then by the property of σ sketched above it follows that $T\ulcorner\phi\urcorner \in \Xi_M^{n+1}(S)$ iff $T\ulcorner\sigma(\phi)\urcorner \in \Xi_M^\delta(S')$. But this is by no means guaranteed since $T\ulcorner\phi\urcorner$ is of quotation degree $m + 1$, which implies that it hasn't stabilized at stage $n + 1$.

To illustrate this point consider the case where $S = \{P^m t_1 \dots t_m\}$ and $n = 0$. In this case we wish to construct a function σ , which maps D respecting condition (\dagger) and for which we have that $\phi \in \Xi_M^1(S)$ iff $\sigma(\phi) \in \Xi_M^\delta(S')$ where δ is an arbitrary ordinal greater than or equal to 1. Suppose furthermore that $M \not\models P^n t_1 \dots t_m$ and define $\sigma(P^n t_1 \dots t_m) = (t_1 \neq t_1)$. Then $P^n t_1 \dots t_m \notin \Xi_M^1(S)$ iff $\sigma(P^n t_1 \dots t_m) \notin \Xi_M^\delta(S')$ but supposing that σ respects condition (\dagger) we have

$$T\ulcorner P^m t_1 \dots t_m \urcorner \in \Xi_M^1(S) \Leftrightarrow T\ulcorner t_1 \neq t_1 \urcorner \in \Xi_M^\delta(S')$$

but $T\ulcorner P^n t_1 \dots t_n \urcorner \in \Xi_M^1(S)$ where for no δ with $1 \leq \delta$, $T\ulcorner t_1 \neq t_1 \urcorner \in \Xi_M^\delta(S')$. That is, σ does not respect the interpretation of the truth predicate after all.

Moreover, intuitively it is pretty clear why the lemma can't go through given the present definitions once we have a predicate $true(x, y)$. Consider the sentence

$\forall x(true(x, \int\phi\int) \rightarrow Tx)$. According to our definitions this sentence has the quotation degree of $\int\phi\int$, but clearly whether it is true or false depends on the truth of the sentence $T\int T\int\phi\int$ and hence of a sentence of $qd(\int\phi\int) + 1$. So, in this case we would expect the sentence to stabilize at stage $n + 3$.²⁹

Using Lemma 2.7 we can establish our main claim namely that we reach a fixed-point at the first limit ordinal ω which implies that for any premodel M with domain D and $S \subseteq D$ the \mathcal{L}_{QT} -model $(M, \Xi_M^\omega(S))$ satisfies (TB):

Theorem 2.10 *For any $S \subseteq D$ and any premodel M and all sentences ϕ*

- (i) $\phi \in \Xi_M^\omega(S) \Leftrightarrow (M, \Xi_M^\omega(S)) \models \phi$
- (ii) $(M, \Xi_M^\omega(S)) \models T\int\phi\int \Leftrightarrow \phi$

Theorem 2.10 may be read as establishing that, contra Tarski, Convention T can be satisfied in a semantically closed language.

At this point we do not wish to evaluate whether this construction might serve as a viable conception of truth in its own right. However, what seems important is to realize that the construction can be generalized in order to rebut Montague's assessment that virtually all of modal logic must be sacrificed, if we treat modalities syntactically—at least, if this assessment is understood in its straightforward, general way. To some extent this had been already shown to be false by Niemi, but Niemi's account focuses on one particular modal system and one single modality.

2.2.2 Models for Modalities Conceived as Predicates

We now generalize Gupta's account and introduce an arbitrary finite number of modal notions into the picture. To this end we combine the construction we have just sketched with possible world semantics as known from modal operator logic. The strategy of employing some form of possible world semantics to truth-theoretic constructions to obtain an interpretation for the modal predicates will play an important role throughout our investigation and the present construction will thus prove instructive.

Indeed, we use the fact that possible world semantics for modal operator logic and the semantics to be constructed display a similar structure for showing that consistent accounts of modality are not only possible but adequate from the perspective of modal operator logic. In that we follow the lead of Skyrms [152] and Schweizer [148] who already established—using similar techniques—the adequacy of syntactical modal logic for a single modality with respect to the modal operator logic $S5$.

Extending the approach to multiple modalities is interesting from two perspectives. First, natural language incorporates multiple modalities and thus a formal

²⁹Obviously, using the predicate $true(x, y)$ repeatedly we can increase the quotation degree and the stage at which the sentence is expected to stabilize even further.

approach allowing for the joint treatment of several modal notions seems to be interesting from this perspective. But second, Niebergall [116], Halbach [64, 65] and, Horsten and Leitgeb [77] have shown that new inconsistency results might arise, if we allow for multiple, interacting modal predicates. Moreover, as a bonus, the multimodal framework to be sketched provides an outline of how the construction we lay out in the last chapter of our work can be generalized to a multimodal setting.

The basic idea of the present proposal is to construct a possible world semantics in quantifying over “Gupta-style” models. Consequently, we will evaluate a modal predicate with respect to a class of “accessible” models. More precisely, the interpretation of a modal predicate will be the intersection of the interpretations of the truth predicates in the models accessible from the present model.

In “A Syntactical Approach to Modality” Schweizer [148] puts forward a similar approach for he also quantifies over “Gupta-style” models. But his approach is less general in two respects. First, Schweizer considers only one modality and therefore does not provide an account that can deal with multiple modalities, but second, Schweizer only considers the modal logic S5 and thus does not allow the accessible models to vary from one model to another.³⁰

A related account has equally been developed by Asher and Kamp [10, 11] who also construct a possible world semantics based on Gupta's and Herzberger's Revision Theory of Truth. The work by Asher and Kamp [11] is probably closest to the present undertaking, however, they work in a slightly different setting. In their paper they only consider one modal predicate, but a generalization to multiple modal predicates seems to be rather straightforward.

We will work in a language \mathcal{L}_{QM} , which is like \mathcal{L}_{QT} except that we add a finite number of one place modal predicates, say N_1, \dots, N_n . The expressions ‘term’, ‘formula’, and ‘quotation degree’ are defined as in Definition 2.3. We do not reproduce the definition. Accordingly, our language possesses quotation names for all the sentences of the language. We may also adopt Definition 2.4 without change.

Contrary to the case of a single truth predicate we discussed above, the interpretation of truth and the modal predicates will not simply be defined relative to one model but rather the interpretations will depend upon a modal frame. A modal frame consists of a nonempty set of premodels \mathcal{W} , which will serve as our set of possible worlds and a finite number of accessibility relations defined on this set. Besides the notion of a modal premodel frame we need the notion of an evaluation. An evaluation assigns to every world, i.e., every premodel in \mathcal{W} , a set of sentences of \mathcal{L}_{QM} which will serve as the interpretation of the truth predicate at this world³¹:

Definition 2.11 (*Modal Premodel Frame, evaluation function*) Let $\mathcal{W} \neq \emptyset$ be some set of premodels, i.e., $\mathcal{W} \subseteq \mathfrak{M}$, and R_1, \dots, R_n dyadic “accessibility” relations

³⁰In possible world semantics for modal operator logic the logic S5 characterizes a frame based on an equivalence relation. It can be shown that in this case the accessibility relation can be dropped instead of quantification *simpliciter* over the set of possible worlds. Cf. Hughes and Cresswell [78] and Blackburn et al. [15] for more on modal logic and possible world semantics.

³¹Since the domains of the premodels may vary we decided to avoid complication and to consider subsets of the set of sentences of \mathcal{L}_{QM} as interpretations of the truth predicate only.

on \mathcal{W} . Then $F = \langle \mathcal{W}, R_1, \dots, R_n \rangle$ is called a modal premodel frame. A function $f : \mathcal{W} \rightarrow P(\text{Sent}_{\mathcal{L}_{QM}})$ is called an evaluation function relative to F . The set of all evaluation functions relative to a frame F is denoted by Val_F .

A modal premodel frame together with an evaluation function induces a model for \mathcal{L}_{QM} relative to each member of \mathcal{W} .

Definition 2.12 (*Models for \mathcal{L}_{QM}*) Let F be a frame and $f \in \text{Val}_F$ an evaluation function. Then F and f induce a model $M^f = (M, f(M), Y_M^1, \dots, Y_M^n)$ of the language \mathcal{L}_{QM} relative to every premodel (world) M . $f(M)$ is the extension of the truth predicate at M and Y_M^i with

$$Y_{Mf}^i = \bigcap_{M' \in [MR_i]} f(M')$$

the extension of the modal predicates.³² $[MR_i]$ is short for the set $\{M' \in \mathcal{W} : MR_i M'\}$.

Definition 2.13 (*Modal Jump relative to a premodel frame*) Let Val_F be the set of evaluation functions of a modal premodel frame F . The modal jump relative to F , Ξ_F , is an operation on Val_F such that for all $f \in \text{Val}_F$ and all $M \in \mathcal{W}$

$$[\Xi_F(f)](M) = \{\phi \in \text{Sent}_{\mathcal{L}_{QM}} : M^f \models \phi\}$$

Iterative applications of Ξ_F for a given ordinal α are defined by transfinite recursion as in Definition 2.6:

$$\Xi_F^\alpha(f) := \begin{cases} f, & \text{if } \alpha = 0 \\ \Xi_F(\Xi_F^\delta(f)), & \text{if } \alpha = \delta + 1 \\ g \in \text{Val}_F, & \text{if } \alpha \text{ is a limit ordinal} \end{cases}$$

where for all $M \in \mathcal{W}$

$$g(M) = \{\phi \in \text{Sent}_{\mathcal{L}_{QM}} : \exists \kappa (\kappa \leq \beta < \alpha \text{ and } [\Xi_F^\kappa(f)](M) \models \phi)\}.$$

Given a frame F we sometimes write f^α instead of $\Xi_F^\alpha(f)$.

Thus the iterative application of the modal jump relative to a frame leads us to a sequence of interpretations of the truth predicate and the modal predicates. As before it can be shown that the process reaches a unique fixed-point at the first limit ordinal ω , i.e., Lemma 2.7 carries over with minor modification.

Lemma 2.14 *Let F be a modal premodel frame and $f, g \in \text{Val}_F$. Then for all $M \in \mathcal{W}$, all natural numbers n and all ordinals α , if $\alpha > n + 1$ then for all sentences ϕ with $dq(\phi) \leq n$:*

³²We assume \bigcap to be an operation on $P(\text{Sent}_{\mathcal{L}_{QM}})$ and thus, in particular, $\bigcap \emptyset = \text{Sent}_{\mathcal{L}_{QM}}$.

$$\phi \in [\Xi_F^{n+2}(f)](M) \Leftrightarrow \phi \in [\Xi_F^\alpha(g)](M)$$

Proof Sketch We use the construction of the proof of Theorem 2.7 and observe that we can construct the mapping σ such that it respects the interpretation of the truth and the modal predicates, i.e., $\sigma : D \rightarrow D$ is a bijection which is the identity function on the sentences of quotation degree $< n$ and on $D - \text{Sent}_{\mathcal{L}_{QM}}$ and additionally,

$$\begin{aligned} \phi \in [\Xi_F^{n+1}(f)](M) &\Leftrightarrow \sigma(\phi) \in [\Xi_F^\delta(g)](M) \\ \phi \in Y_{M^{f^{n+1}}}^i &\Leftrightarrow \sigma(\phi) \in Y_{M^g}^i \end{aligned}$$

for all i , with $1 \leq i \leq n$. We can find such a function σ due to the following observations: let U be the interpretation of a modal predicate at a stage γ and V the interpretation of either the truth predicate or one of the remaining modal predicates, then either

- U and V coincide, or
- U is a denumerable subset of V , where $V - U$ is also denumerable, or
- $U \cap V$ is denumerable set of sentences, and so are $U - V$ and $V - U$.³³

Then given such a function σ the proof goes through as before. \square

Corollary 2.15 *Let F be a frame. Then for all evaluation functions $f, g \in \text{Val}_F$ and all $\phi \in \text{Sent}_{\mathcal{L}_{QM}}$*

- (i) $\Xi_F^\omega(f) = \Xi_F^{\omega+1}(g)$
- (ii) $M^{f^\omega} \models \phi \Leftrightarrow \phi \in f^\omega(M)$

The corollary implies the existence of unique fixed-points of Ξ_F in Val_F , i.e., there exists exactly one evaluation function $g \in \text{Val}_F$ with

$$\Xi_F(g) = g$$

This allows us to define the notion of a proper model of \mathcal{L}_{QM} :

Definition 2.16 (*Proper Model*) Let F be a frame, g the evaluation function in Val_F with $\Xi_F(g) = g$ and $M \in \mathcal{W}$ a proper premodel. Then the model M^g induced by F and g relative to M is called a proper model of \mathcal{L}_{QM} . We write $F, M \models \phi$ to convey the fact that ϕ is true in the proper model induced by the premodel M relative to

³³To see this, note that the interpretations of the truth and modal predicates cannot be disjoint since all the tautologies and the \mathcal{L} -truths are in the interpretation of the truth predicate in every model. If there exists a sentence which is not in the interpretation of the one predicate but the other, then there exist denumerable many sentences of this kind: take all the conjunctions of tautologies with this sentence. This explains the second observation. A parallel argument establishes the third.

F .³⁴ If a formula $\phi \in \mathcal{L}_{QM}$ is true in all proper models induced by a frame, we write $F \models \phi$. Let \mathfrak{F} be a class of frames, if for all $F \in \mathfrak{F}$ ($F \models \phi$) we write $\mathfrak{F} \models \phi$.

(TB) is true in any modal premodel frame. Moreover, in virtue of our construction the modal principle (K) holds for every modal predicate, i.e.,:

Theorem 2.17 *Let \mathfrak{F} be the class of all modal premodel frames. Then for all $\phi, \psi \in \text{Sent}_{\mathcal{L}_{QM}}$*

$$\begin{aligned}\mathfrak{F} &\models T\mathfrak{J}\phi\mathfrak{I} \leftrightarrow \phi \\ \mathfrak{F} &\models N_i\mathfrak{J}\phi \rightarrow \psi\mathfrak{I} \rightarrow (N_i\mathfrak{J}\phi\mathfrak{I} \rightarrow N_i\mathfrak{J}\psi\mathfrak{I})\end{aligned}$$

for all i with $1 \leq i \leq n$. Moreover, if ϕ is true in all proper models induced by a frame F , then $N\mathfrak{J}\phi\mathfrak{I}$ will be true in all models induced by F .

At first sight Theorem 2.17 might seem disappointing since our models only satisfy the modal principle (K) where, without doubt, we wish further modal principles to be true. However, this is just parallel to the situation in possible world semantics for modal operator logics. If no assumptions are made on behalf of the accessibility relations of a modal frame, the models based on this frame are only guaranteed to validate the operator versions of (K).

But as in possible world semantics for modal operator logic we can impose conditions on the accessibility relation R_1, \dots, R_n of a modal premodel frame in order to ensure that further modal principles will be true in a proper model. For example, if we require an accessibility relation R_i of a modal premodel frame F to be reflexive, then for every proper model \mathcal{M} induced by F the (T) principle will turn out true:

$$\mathcal{M} \models N_i\mathfrak{J}\phi\mathfrak{I} \rightarrow \phi.$$

Basically, this can be done for all the common, characteristic modal principles including those principles that deal with the interaction of the modal predicates. This highlights again how the present semantic approach parallels possible world semantics for modal operator logic to a certain extent. Indeed, we may impose one and the same condition on the accessibility relation of a modal premodel frame and on the accessibility relation of a possible world frame for modal operator logic.

Definition 2.18 (*Property Φ*) A modal premodel frame and a modal frame of possible world semantics for modal operator logic can share the conditions imposed on the accessibility relation. We call these conditions “property Φ ”.

Comparing the effects of imposing certain properties on the accessibility relation shows that the two semantics work pretty much in the same way. In general, if a certain property Φ is imposed on the accessibility relations of a modal premodel frame and a modal frame for modal operator logic respectively, the “same” modal principles will

³⁴Note that as a consequence of the uniqueness of fixed-points for every frame F and proper premodel $M \in \mathcal{W}$ there exists only one proper model, which is induced by M relative to this frame.

be true in both frames. To use our example once more: if the accessibility relation is reflexive, the predicate version and the operator version of the modal principle (T) will be true in the modal premodel frame or, respectively, the modal frame, i.e., the possible world frame of modal operator logic.

We can exploit the structural similarities between the two semantics in order to show that the syntactical approach to modalities is adequate with respect to a wide class of multimodal operator logics. Moreover, the semantics that we have developed is rather flexible and intuitive to the point possible world semantics for modal operator logic is. Accordingly, especially the champion of possible world semantics should be tempted by the present approach.

2.2.3 Adequacy of the Predicate Approach to Modalities

The idea behind the adequacy result is to show that the syntactical approach validates exactly the theorems of modal operator logic. Of course, it remains to be specified what it means for the syntactical approach to validate the same theorems, as we are dealing with two different languages and, as a trivial consequence, a theorem of modal operator logic cannot be true in the syntactical approach (nor can it be false).

Still the gist of the affirmation seems to be pretty clear. If a modal operator logic proves a certain modal statement, then the corresponding modal statement of the quotation name language, i.e., the statement in which the modal operators of the initial statement are now read as modal predicates and the formula in the scope of the operator will be placed between quotation marks in order to form a quotation name, should be true in an appropriate modal premodel frame. The idea is to translate the propositional atoms of the modal operator language into sentences of the language \mathcal{L}_{QM} and to guarantee that the translation function, say H , commutes with the boolean connectives and translates a formula $\Box_i \phi$ as $N_i \int H(\phi) \int$. To this end let $\mathcal{L}_{\blacksquare \Box_i}$ be a multimodal language in the sense of Definition 1.3 where \blacksquare is meant to be a truth operator. We define the notion of a translation H^* over a realization $*$.

Definition 2.19 (*Translation*) A mapping $*$: $At_{\blacksquare \Box_i} \rightarrow Sent_{\mathcal{L}_{QM}}$ is called a realization. H^* is a translation function from $\mathcal{L}_{\blacksquare \Box_i}$ iff it respects the following conditions:

$$H^*(\phi) := \begin{cases} \phi^*, & \text{if } \phi \in At_{\blacksquare \Box_i} \\ \perp, & \text{if } \phi \doteq \perp \\ \neg H^*(\psi), & \text{if } \phi \doteq (\neg \psi) \\ H^*(\psi) \wedge H^*(\chi), & \text{if } \phi \doteq (\psi \wedge \chi) \\ T \int H^*(\psi) \int, & \text{if } \phi \doteq (\blacksquare \psi) \\ N_i \int H^*(\psi) \int, & \text{if } \phi \doteq (\Box_i \psi) \end{cases}$$

Using this notion of translation we can spell out our adequacy condition more precisely. A sentence of the modal operator language is a theorem of a modal operator

logic under consideration, if and only if, for every realization its translation will be true in an appropriate class of modal premodel frames of the syntactical approach.

The class of modal premodel frames appropriate with respect to a modal operator logic will be determined with respect to property Φ : if a modal operator logic is complete with respect to the class of modal frames with property Φ , the class of modal premodel frames with property Φ will be considered as appropriate.

Theorem 2.20 *Let $\mathcal{F}_{\blacksquare\Box}$ be the class of possible world frames with property Φ and \mathfrak{F} the class of modal premodel frames with property Φ . Then for all $\phi \in \mathcal{L}_{\blacksquare\Box_i}$*

$$\mathcal{F}_{\blacksquare\Box} \models \phi \Leftrightarrow \text{for all realizations } * (\mathfrak{F} \models H^*(\phi)).$$

To establish Theorem 2.20 we employ two lemmata. The first one shows us how to construct a possible world frame starting from a proper premodel frame whereas the second establishes the converse direction:

Lemma 2.21 *For all modal premodel frames F with property Φ , there exists a possible world frame $F_{\blacksquare\Box}$ with property Φ such that for all $\phi \in \mathcal{L}_{\blacksquare\Box_i}$:*

$$F_{\blacksquare\Box} \models \phi \Leftrightarrow \text{for all realizations } * (F \models H^*(\phi)).$$

Proof Let $F = \langle \mathcal{W}, R_1, \dots, R_n \rangle$ be the modal premodel frame. Now, take \mathcal{W} to be the set W of worlds of the possible world frame of modal operator logic and define $R_{\Box_i} := R_i$. Set $F_{\blacksquare\Box} := \langle W, R_{\Box_1}, \dots, R_{\Box_n} \rangle$. We may verify by an induction over the complexity of ϕ that for all $M \in \mathcal{W}$

$$F_{\blacksquare\Box}, M \models \phi \Leftrightarrow \forall * (F, M \models H^*(\phi)).$$

We discuss $\phi \doteq \Box_i \psi$ and leave the rest to the reader. To this end assume $F_{\blacksquare\Box_i}, M \models \Box_i \psi$. By Definition 1.6 the latter is equivalent to

$$\text{for all valuations } V \forall M' (M R_{\Box_i} M' \Rightarrow (F_{\blacksquare\Box_i}, V), M' \models \psi).$$

Since V does not occur in the antecedent this is equivalent to

$$\forall M' (M R_{\Box_i} M' \Rightarrow \forall V ((F_{\blacksquare\Box_i}, V), M' \models \psi)).$$

By induction hypothesis and definition we get

$$\forall M' (M R_i M' \Rightarrow \forall * (F, M' \models H^*(\psi)))$$

and since $*$ is not bound in the antecedent of the above conditional we may conclude to the desired. \square

Lemma 2.22 *For all possible world frames $F_{\blacksquare\Box}$ with property Φ , there exists a modal premodel frame F with property Φ such that for all $\phi \in \mathcal{L}_{\blacksquare\Box_i}$:*

$$F_{\blacksquare\Box} \models \phi \Leftrightarrow \text{for all realizations } * (F \models H^*(\phi)).$$

Proof Let $F_{\blacksquare\Box}$ be the frame $\langle W, R_{\Box_1}, \dots, R_{\Box_n} \rangle$. Take some set of proper premodels A with $|W| \leq |A|$ and let $\gamma : W \rightarrow A$ be some injective function. We define

$$\begin{aligned} \mathcal{W} &:= \{\gamma(w) : w \in W\} \\ R_i &= \{(\gamma(w), \gamma(v)) : w R_{\Box_i} v\} \end{aligned}$$

and set $F = \langle \mathcal{W}, R_1, \dots, R_n \rangle$. We may then verify by induction on the complexity of ϕ that

$$F_{\blacksquare\Box}, w \models \phi \Leftrightarrow \forall * (F, \gamma(M) \models H^*(\phi))$$

which establishes the lemma. \square

We may now state the proof of the main theorem as a trivial consequence of the two lemmata:

Proof of Theorem 2.20 Suppose $\mathcal{F}_{\blacksquare\Box} \models \phi$ but that there exists a modal premodel frame F with property Φ and $\neg \forall * (F \models H^*(\phi))$. Then by Lemma 2.21 we end up in contradiction. On the other hand assuming $\forall * (\mathfrak{F} \models H^*(\phi))$ and $\mathcal{F}_{\blacksquare\Box} \not\models \phi$ leads to contradiction by Lemma 2.22. \square

We get a nice corollary from this result which can be read as the exact formal rendering of the adequacy criterium we laid out:

Corollary 2.23 *Let \mathcal{S} be a modal operator logic complete with respect to the class of possible world frames \mathcal{F} with property Φ . Let \mathfrak{F} be the class of modal premodel frames with property Φ , then for all $\phi \in \mathcal{L}_{\blacksquare\Box_i}$:*

$$\mathcal{S} \vdash \phi \Leftrightarrow \forall * (\mathfrak{F} \models H^*(\phi))$$

Proof By completeness of \mathcal{S} we have for all $\phi \in \mathcal{L}_{\blacksquare\Box_i}$

$$\mathcal{S} \vdash \phi \Leftrightarrow \mathcal{F} \models \phi$$

which together with Theorem 2.20 establishes the claim. \square

In the light of these results we to conclude that:

- The unrestricted characteristic modal principles can be consistently maintained in the predicate setting. Moreover, this claim also holds for the modal principles characterizing the interaction of multiple modalities.
- Predicate treatments of modality are adequate with respect to modal operator logic in the sense outlined above.

2.3 Operators and Predicates

The syntactical approach to modality we have just outlined appears to vindicate predicate approaches to modality from the perspective of modal operator logic. As we have seen, the approach allows us to adhere to the unrestricted modal principles and also proves to be adequate from the perspective of modal operator logic. Moreover, the approach comes with a semantics, which is intuitive to the extent possible world semantics for the modal operator can be considered as intuitive. And even though the approach we have outlined so far does not come with a developed proof theory the result of Asher and Kamp [11], who provide a complete axiomatization of their semantics in the absence of vicious forms of self-reference, suggest that this can be done along the lines of Niemi [117].

A common reaction to vindications of the predicate approach that embrace the unrestricted modal principles, which in other settings would lead to paradox, has been to question whether these approaches can be considered as predicate approaches to modality proper. The argument which appears at least implicitly in the literature can be glossed as follows: the fact that you have shown that modal logic, i.e., the principles of modal operator logic, can be upheld in a predicate setting just shows that the modal predicates under consideration are just operators in disguise. This kind of argument is maybe most explicitly stated in the writings of Reinhard [139] and Grim [52].

On the face of it this claim seems puzzling as ultimately the only precise way to distinguish an operator from a predicate is by the formation rules of the language under consideration: if the expression applied to a sentence yields a new sentence, it is an operator but if the expression needs to be applied to a name or term to yield a sentence, it is a (one-place) predicate. But the modal expressions of the approach sketched in the preceding section apply to terms to form sentences and thus the modal expressions are predicates.

Now, we have already seen that this distinction is more blurred than one might initially think. In Skyrms' [152] syntactical approach to modality the construction rule for a modal formula is given by the following clause:

- If S is a sentence of L_n , then S and $*Q(S)$ are sentences of L_{n+1} .

where the $*$ is supposed to be the modal predicate and $Q(S)$ a quotation name of S . Additionally, L_{n+1} is closed off under propositional connectives. L_n is the language previously constructed.³⁵ But as we have pointed out before the quotation names are not ordinary terms since only the modal predicate $*$ is applicable to them. They only appear in the formation rule quoted above. Similarly, the modal predicate can only be applied to these quotation names and not to arbitrary terms of the language. Therefore, one is tempted to conclude that Skyrms' modal predicate is not a predicate

³⁵ Ultimately the modal language L is obtained by considering the union of all these languages, that is $L := \bigcup_{n \in \omega} L_n$.

proper, but an modal operator in disguise, as the formation rule of his predicate is just the formation rule of the modal operator \Box in a more convoluted way:

- If S is a sentence of L_n , then S and $\Box S$ are sentences of L_{n+1} .³⁶

It seems as if Skyrms reads some quotational mechanism into the modal operator whereas this quotational mechanism is benign to the modality under consideration, i.e., it does not interact with the rest of the language. This might then be, philosophically speaking, an interesting thesis, but it fails to be of any interest from the formal point of view, as we can without consequences dispense of the quotation name of a sentence S , $Q(S)$, in favor of the modal operator $*(Q(\bullet))$. That is, if we read $*(Q(\bullet))$ not as two expression with an argument position each, but as one expression with one argument position we're back in the operator setting.

So there is a point in claiming that Skyrms' predicate is not a predicate proper but an operator in disguise or, more precisely, abstracted from the modal operator by introducing a superfluous (though maybe philosophically justifiable) quotation or respectively abstraction mechanism. When the names of the sentences are taken to be their arithmetical names, the situation becomes a bit more blurred since these names can of course appear in argument positions of other predicates or function symbols. Yet only the numerals, which are names of formulas occurring in the construction process, can appear in the argument position of the modal predicate. We have already argued that this approach has its drawbacks but to its credit it does not seem as artificial as the quotation name approach advocated by Skyrms for the terms appearing in modal formulas are proper terms of the language. Still, the construction rule for modal formulas is mimicked after the one for the modal operator and variables are not permitted in the argument position of the modal predicate.

A slight variant of this latter proposal, put forward by Belnap and Gupta [14], allows all numerals to appear in the argument position of the modal predicate but no other terms or variables. In this setting the standard laws of modal logic can be upheld likewise. However, it seems that the formation rule governing the behavior of the modal predicate is no longer a copy of the formation rule governing the modal operator. Is this predicate then also an operator in disguise?

Well, in a certain way the predicate behaves like an operator as Belnap and Gupta show. They show (cf. [14], pp. 239–242) that the modal predicate language is inter-translatable with respect to the modal operator language in an intended way.³⁷ Given two translation functions which translate the modal predicate into the modal operator and *vice versa*, versions of Theorem 2.20 can be proven for both directions. That is, it can be proven that if a formula ϕ of the modal predicate language is valid in an intended semantics for this very language, then so is its translation in an intended semantics for the modal operator logic, and vice versa.

³⁶See again Skyrms [152], p. 369.

³⁷The two languages only differ with respect to their treatment of modality, i.e., both are languages of arithmetic where the one is supplemented by a modal operator but the other by a modal predicate.

This might be taken to show that the modal predicate language is just a notational variant of the modal operator language³⁸ and thus the modal predicate just an operator in disguise. But despite the fact that the two languages allow us to say the same and are thus notational variants, the modal predicate is, syntactically speaking, a predicate and not an operator. However, it is a predicate which is severely crippled by the formation rules of the language and one might thus argue that for any approach to modality to be a syntactical or predicate approach to modality proper, the modal predicates should be predicates on which no syntactic restrictions have been placed.

In less abstract terms this should come down to the requirement that the modal predicates should behave like ordinary predicates in that they allow for quantification into their argument position and they should be open to diagonalization, if the necessary background assumptions are met.

One way to cash out these ideas is to require modal predicates to be proper predicates in the sense defined below

Definition 2.24 (*Proper predicate*) Let \mathcal{L} be an arbitrary first-order language. We call an \mathcal{L} -predicate P^n , $n \in \omega$, a proper predicate with respect to \mathcal{L} . If all the three following conditions are satisfied in \mathcal{L}

1. for all $t_1, \dots, t_n \in \text{Term}_{\mathcal{L}}$, $P^n t_{j_1}, \dots, t_{j_n}$ is a well formed formula of \mathcal{L} ;
2. for all function symbols f^j , $j \in \omega$ and all $t_1, \dots, t_j \in \text{Term}_{\mathcal{L}}$ $f^j(t_1, \dots, t_j)$ is a well formed term of \mathcal{L} .
3. for all \mathcal{L} -predicates Q^m (including identity), $m \in \omega$ and all $t_1, \dots, t_m \in \text{Term}_{\mathcal{L}}$, $Q^m t_1, \dots, t_m$ is a well formed formula.

Otherwise a predicate is called improper.³⁹

The idea behind the definition of a proper predicate is straightforward. The predicate should be applicable to all terms of the language, and the remaining predicates of the language should be applicable to all the terms the target predicate is applicable to and similarly for function symbols. The definition imposes very restrictive conditions on proper predicates. For example, let $\mathcal{L}_{\in, T}$ be the language of set theory extended by a typed truth predicate, for which the typing restriction is strong in the sense that the formation rules of the language do not allow to construct formulas in which the truth predicate is applied to names of sentences in which the truth predicate occurs, then \in does not count as a proper predicate because not all terms, which are \in -applicable are equally T -applicable. Conversely, but maybe less surprisingly T is not a proper predicate of $\mathcal{L}_{\in, T}$.

This shows that a predicate which Tarski would have labeled a “semantical” predicate can be proper only, if the language is (semantically) closed. In fact, the

³⁸We owe this formulation to Belnap and Gupta:

Despite the differences in logical grammar of necessity, \mathcal{PL}^- [the modal predicate language, J.S.] and \mathcal{CL} [the modal operator language, J.S.] are *notational variants* [our italics]. (cf. [14], p. 242.)

³⁹Clearly, condition 3 implies condition 1, we have nonetheless included condition 1 in our definition in order to highlight the fact that a proper predicate needs to be applicable to all terms of the language.

condition imposed by Definition 2.24 is stronger than the one of a semantically closed language. That is, if a predicate is proper in \mathcal{L} , then \mathcal{L} is semantically closed, however, the converse need not hold. Skyrms' construction provides a counterexample. His target language L_ω possesses names for all the sentences of the language but clearly his predicate $*$ is not proper. In fact neither of condition (1), (2) and (3) of Definition 2.24 are met.

If we require the modal predicate to be a proper predicate, this also rules out, as expected, the proposal of Belnap and Gupta. In the light of discussion one might think that the notion of a proper predicate as defined in Definition 2.24 can be used to vindicate Montague. That is, once we require the modal predicates to be proper predicates then virtually all of modal logic has to be given up. We already know that this can't be right. The quotation name approach we sketched out in some detail shows this claim to be false since, maybe surprisingly, the modal predicates of this approach are proper predicates in the sense of Definition 2.24.

Does this show that the quotation name approach we have outlined is an impeccable predicate approach to modality and Montague's claim just false? Are the complaints that the modal predicates are just operators in disguise utterly pointless in this case? From a purely syntactic point of view this is certainly right, the modal predicate satisfies all conditions one can reasonably demand an expression to fulfill in order to count as a predicate. Most importantly, the predicate applies to ordinary terms, allows quantification into its argument position and can be diagonalized.

All this suggests that the talk of modal predicates being operators in disguise is actually beside the point, as from a syntactic point of view it is not clear why the syntactical predicates should be considered as operators in disguise. A more interesting question seems to be whether a modal predicate allows us to say anything more than we could using the modal operator. For example, we already saw that Belnap and Gupta's modal predicate does—in some relevant sense—not allow us to do so. Accordingly, rather than questioning whether the modal predicates of the quotation name approach—or any other predicate approach to modality embracing the standard laws of modal logic—are predicates or operators in disguise one might query whether the approach takes us beyond the operator approach. Theorem 2.20 might be interpreted as saying that the operator approach and the quotation name approach agree on the propositional modal fragment of the quotation name language. But there is a decisive difference between the two approaches: the predicate approach allows quantification into the argument position of the modality where nothing of the like is possible in the operator setting as it is usually conceived.

2.3.1 *Predicates and Quantification*

If we conceive of modal notions as predicates, we can by appeal to standard first-order quantification quantify into the argument position of the modal notion at stake and

thereby provide formalization of universal or existential claims.⁴⁰ This feature of the predicate approach to modality is a distinguishing feature of the predicate approach to modality and is often seen as one of the main advantages of the predicate approach. Accordingly, any predicate approach to modality will be judged relative to its capacity of formulating quantified principles and, for instance, Skyrms' proposal fails badly in this respect, as it does not allow for quantification into the argument position of the modal notion.

In contrast, the quotation name approach allows us to quantify into the argument position of the modal and the truth predicate which in turn makes it possible to formulate quantified principles of certain modal principles. For example, we can express a quantified version of the factivity principle in our framework, i.e.,

$$(T_Q) \quad \forall x (Sent_Q(x) \rightarrow (Nx \rightarrow Tx)).$$

Still, the quotation name approach is rather disappointing with respect to its capacity of formulating quantified principles. On the face of it, we cannot even state generalizations of key modal principles like (K) . To appreciate this notice that

$$(BK) \quad \forall x, y (Sent_Q(\int x \rightarrow y) \rightarrow (N\int x \rightarrow y \rightarrow (Nx \rightarrow Ny)))$$

is not a well-formed formula of \mathcal{L}_{QM} and even if it were, the variable 'x' would not occur in the antecedent of (BK) , but only its name and thus (BK) would fail to reflect the generalization of modal principle (K)

$$(K_{SQ}) \quad N\int \phi \rightarrow \psi \rightarrow (N\int \phi \rightarrow N\int \psi).$$

The problem is that we need to quantify into quotation contexts which at least *prima facie* is not possible. In order to do so, we need to know that $\int \phi \rightarrow \psi$ is the name of a sentence whose main connective is material implication. In other words we need information about the structure of the quotation name, that is, we need to be able to speak about syntax in the language itself.

In Remark 2.8 we have already seen that we can introduce predicates for all truth functional connectives and still maintain the unrestricted principle (TB) in the case of truth and this observation carries over to the modal case. For instance, we can introduce a three place predicate $Imp(x, y, z)$ saying that x is the sentence that y materially implies z and adhere to the unrestricted (schematic) modal principles. Once we have such a predicate at our disposal we can spell out a general version of (K) as follows:

$$(K_Q) \quad \forall x, y, z (Sent_Q(y) \wedge Sent_Q(z) \rightarrow (Imp(x, y, z) \wedge Nx \rightarrow (Ny \rightarrow Nz)))$$

⁴⁰By this we mean that the argument position of the modal notion is a term position which can be occupied by a variable and thus can be bound by quantifiers. This is to be distinguished from "quantifying in" which is often understood as quantifying into a formula which occupies the argument position of a modal operator.

Further quantified principles can be spelled out, if we allow further predicates expressing concepts of syntax as negation $Neg(x, y)$, conjunction $Con(x, y, z)$ or disjunction $Dis(x, y, z)$. For example, by means of the predicate for negation we can state a general version of the consistency principle (Con), that is:

$$(Con_Q) \quad \forall x, y (Sent_Q(y) \rightarrow (Nx \wedge Neg(x, y) \rightarrow \neg Ny))$$

However, as we have already hinted at in Remark 2.9, we cannot hope to introduce concepts of syntax *ad libitum*: if “enough” theory of syntax is available, we may define a substitution function, which, as we shall discuss later, together with the unrestricted Tarski biconditionals or, alternatively, certain unrestricted modal principles proves sufficient for deriving the paradoxes. Moreover, the fact that we cannot allow ourselves to all concepts of syntax we may hope for can also be seen as a corollary of Quine’s observation [129] that Peano Arithmetic can be interpreted in the theory of syntax. Consequently, a version of the Diagonal lemma may be proven in this very theory such that the names appearing as names of the sentences in this version of the Diagonal lemma would be the quotation names themselves and thus Montague’s theorem would apply.

It remains an interesting question to see how much syntax one can exactly introduce in the presence of the unrestricted modal principles without paradox to arise. As we have argued in Remark 2.9, it remains an open question whether we can introduce a predicate saying that a sentence is the T -necessitation of another sentence and this observation carries over to predicates saying that a sentence is the N -necessitation of another sentence. If we had such predicates at our disposal, we could spell out general principles involving iterated modalities. For example, as it stands we can not spell out a quantified version of the modal principle

$$(4_{SQ}) \quad N\}\phi\} \rightarrow N\}N\}\phi\}\}$$

within the quotation name approach. In the presence of a predicate $Nec(x, y)$ we could do so in the following way:

$$(4_Q) \quad \forall x, y (Sent_Q(x) \rightarrow (Nx \wedge Nec(y, x) \rightarrow Ny))$$

However, as things are, as soon as we have to deal with iterated modalities the quotation name approach is shut down. And this is a major drawback because, as we mentioned, one of the alleged benefits of a predicate treatment of modalities is to allow for straightforward formalization of general and existential statements involving the modal notions. This drawback seems to be even more telling since it is presumingly quantification which, amongst other things, sets modal operators and predicates apart. Our findings therefore suggest that whereas the modal predicates of the quotation name approach are thus maybe not operators in disguise they do not exhaust the full expressive power of the predicate approach and are in this respect not so much different from the modal operators.

To be sure in a language in which, e.g., we use numerals as the names of the modal sentences, quantified principles involving, e.g., iterated modalities can be spelled out, as—via coding—we can define functions for all the syntactical operations,⁴¹ but of course in this setting certain unrestricted modal principles can no longer be upheld for sake of paradox.

2.3.2 Operators, Quantification, and the Paradoxes of Indirect Discourse

If modalities are expressed by operators, then it is not possible to quantify into argument position of the modal expressions within the boundaries of a first-order modal language. The reason is that the argument position of the modality is occupied by a sentence and we would thus need to quantify into sentence position. However, if the machinery of propositional quantification, i.e., restricted second-order quantification, is introduced, the latter becomes possible and it shows that the introduction of propositional quantifiers allows us to spell out quantified versions of virtually all important modal principles. For example, using propositional quantifiers the modal principle 4 can be represented by the non-schematic quantified formula

$$(Q4) \quad \forall p(\Box p \rightarrow \Box\Box p)$$

and, similarly, other modal quantified statements can be represented using propositional quantification. This suggests that we might be better off sticking to modal operators.

But once we allow propositional quantifiers in the modal language the operator approach is threatened by paradoxes likewise.⁴² These paradoxes, even though closely related to the liar like paradoxes, are paradoxes of indirect discourse and therefore differ from the semantic paradoxes in their canonical presentation. In their simplest variant they follow the outlines of Epimenides' Paradox. The paradoxes have not received as much attention as the paradoxes of direct discourse, that is, the liar like paradoxes, but have been discussed by Prior [124], Thomason,⁴³ Burge [21, 22], Asher [9] and, more recently, by Brandenburger and Keisler [19], Pacuit [118] and, Abramsky and Zvesper [1].⁴⁴

⁴¹However, it is important to note that if we want to speak of syntactical operations proper and not of their codes we need to identify the sentences with their codes. This identification seems rather dubious from a philosophical perspective.

⁴²See also Stern [160] for remarks along these lines.

⁴³Asher [9] as well as Belnap and Gupta [14] allude to a never published manuscript "Paradoxes of Intentionality?" as principle source of inspiration. Very recently joint work of Thomason and Tucker [171] entitled "Paradoxes of Intensionality" where the paradoxes of indirect discourse are discussed was published.

⁴⁴In fact, Brandenburger and Keisler [19] rediscovered and reinvented the paradox of indirect discourse, which in the subsequent literature has been relabeled Brandenburg-Keisler paradox.

We consider the language L_{\ddagger} which is a propositional modal language with one modal operator \ddagger , propositional variables p, p', \dots and propositional quantifier \forall . For expository ease we read the modal operator as ‘Onephrase asserts that’. We set up a hypothetical situation as follows.

- (i) Onephrase asserts that everything Onephrase asserts is not the case.
- (ii) That is the only assertion Onephrase ever makes.

But given this setup (i) can be formalized in L_{\ddagger} by means of propositional quantification as follows

$$(O1) \quad \ddagger \forall p (\ddagger p \rightarrow \neg p)$$

and (ii) gives rise to the following assumption in L_{\ddagger}

$$(O2) \quad \forall p (\ddagger p \rightarrow (p \leftrightarrow \forall p (\ddagger p \rightarrow \neg p)))$$

Assuming (O1), (O2) and the standard logic of quantification we can derive a contradiction:

1. $\forall p (\ddagger p \rightarrow \neg p) \rightarrow (\ddagger \forall p (\ddagger p \rightarrow \neg p) \rightarrow \neg \forall p (\ddagger p \rightarrow \neg p))$ (UI)
2. $\forall p (\ddagger p \rightarrow \neg p) \rightarrow \neg \forall p (\ddagger p \rightarrow \neg p)$ 1, (OI)
3. $\neg \forall p (\ddagger p \rightarrow \neg p)$ 2
4. $\ddagger p \wedge p \rightarrow \forall p (\ddagger p \rightarrow \neg p)$ (O2), (UI)
5. $\exists p (\ddagger p \wedge p) \rightarrow \forall p (\ddagger p \rightarrow \neg p)$ 4, (EG)
6. $\forall p (\ddagger p \rightarrow \neg p)$ 5
7. \perp 3,5.

On the face of it, this conclusion seems very puzzling as we haven’t made any assumption on behalf of the modal operator and simply assumed the ordinary laws of quantification. One might take this to be a vindication of the predicate approach, since in the predicate setting the liar like paradoxes required at least the stipulation of certain modal principles, i.e., not all modal notions were prone to paradoxes in the predicate setting.

But this conclusion might be a bit premature, as there are consistent modal logics with propositional quantification. Whether propositional quantification will lead to inconsistency relies on whether we take the initial, hypothetical scenario to be a possible one which in turn relies on how fine grained we individuate

(Footnote 44 continued)

The Brandenburg-Keisler paradox is essentially the paradox to be presented below. It uses the sentence

Ann believes that Bob assumes that Ann believes that Bob’s assumption is wrong.

which immediately shows to be paradoxical, if we ask the question whether Ann believes Bob’s assumption to be wrong (cf. Brandenburger and Keisler [19], p. 212). Their example, however, omits quantification into the argument position of the modality at stake though uses a reification of a proposition instead, namely, Bob’s assumption. In the modal logic representation of this example, which is due to Pacuit [118], a restricted form of propositional quantification is used however.

propositions—or, more general, the objects of our modal attributions. If the hypothetical scenario is ruled out, we can consistently extend the modal logic under consideration by propositional quantification. Most prominently, the individuation of propositions as sets of possible worlds allows for consistent modal logics with propositional quantification whereas, unsurprisingly, these quantifiers range over sets of possible worlds.⁴⁵ Similarly, approaches taking propositions to be entities *sui generis* and limiting the structural information available with respect to these entities will allow for propositional quantification.⁴⁶

Still, while we might have some quarrels with respect to the above scenario, we should be careful trying to dissolve the paradox by dismissing the hypothetical situation, as more plausible scenarios can be constructed and thus the dismissal has counterintuitive consequences. Asher [9] presents the following example⁴⁷:

Suppose Prior is thinking to himself:

(Pr) Either everything that I am thinking at the present moment is false or everything Tarski will think in the next instant, but not both, is false.

Clearly, if Prior thinks (Pr) to himself at t_0 and Tarski thinks that $2 + 2 = 5$ to himself at t_1 , there will be nothing paradoxical and thus the fact that Prior thinks (Pr) and nothing else to himself does not constitute a problem in this situation. But if Tarski thinks, e.g., that Snow is white to himself at t_1 we end up in paradox. Still it seems counterintuitive to react toward this paradox by stipulating that it is impossible that Prior thinks (Pr) and nothing else at t_0 while Tarski thinks that Snow is white and nothing else at t_1 .⁴⁸ This suggests that we should take the paradoxes of indirect discourse seriously and not try to resolve them by dismissing the hypothetical scenario which we will call—following Asher [9]—a *Prior situation*.

Intuitively, to properly evaluate Prior situations propositions need to be able to refer back to themselves, as this is part of the content of (Pr), of what (Pr) asserts, and thus an adequate individuation of propositions should be capable of expressing self-reference. But if propositions are individuated appropriately in this respect propositional quantification, as argued, will have troublesome consequences. The reason for this is that the propositional quantifier is—and again we concur with Asher [9]—a surrogate of the truth predicate. That is, using propositional quantification we can quantify directly into sentence position and thus generalize over sentences. In a first-order setting this can be done only, if sentential predicates have been introduced into the language, for instance, the truth predicate and we know that in the presence of a sentential predicate like truth care has to be taken in order not to run into the paradoxes of direct discourse. However, by means of propositional quantification we

⁴⁵See for instance Bull [20], Kaplan [80] or Fine [43].

⁴⁶Cf. Thomason [167] for an approach along this line.

⁴⁷The general pattern of the example is apparently due to Jean Buridan but was rediscovered and discussed by Prior in [124].

⁴⁸Moreover, given the temporal ordering this would imply that if Prior thinks (Pr) to himself at t_0 , Tarski cannot think that Snow is white to himself at t_1 which is absurd. For more on this see Prior [124] and Thomason and Tucker [171].

can generalize over sentences without appeal to a truth predicate. For example, we can state a quantified version of the law of excluded middle in the following way:

$$\forall p(p \vee \neg p).$$

Moreover, when we analyze the role of the propositional quantifier in the paradox of indirect discourse, it becomes obvious that we face a similar problem as in the case of the liar paradox. For suppose we try to evaluate whether ‘ $\forall p(\ddagger p \rightarrow \neg p)$ ’ is true. Intuitively, this sentence is true if and only if for all propositions \mathcal{P} if Onephrase asserts that \mathcal{P} then *the* proposition that \mathcal{P} is false. But this seems to depend on whether *the* proposition that $\forall p(\ddagger p \rightarrow \neg p)$ is true unless there has been a proposition \mathcal{P} to falsify $\forall p(\ddagger p \rightarrow \neg p)$. However, if *the* proposition that $\forall p(\ddagger p \rightarrow \neg p)$ were true, we would have found a proposition \mathcal{P} which Onephrase asserts and which is true and ‘ $\forall p(\ddagger p \rightarrow \neg p)$ ’ would be false. Thus it seems as if we have ended in an circle similar to the one we encounter in connection with the liar sentence λ where the truth of $T[\lambda]$ depends on whether λ , that is $\neg T[\lambda]$, is true.

If this analysis is correct, it is not surprising that propositional quantification leads to contradiction provided the structure of the propositions is relevant with respect to their evaluation. Since propositional quantification appears to be a surrogate of the truth predicate, instantiating a universally quantified formula has a similar effect as disquotation in the case of the truth predicate. But we know that in the case of the truth predicate we can’t adhere to an unrestricted principle of disquotation, that is, (TB) if self-referential statements can be formulated, or put differently, if we have “enough” syntax at our disposition.⁴⁹ If the modalities are treated as operators the paradoxes of indirect discourse seem to suggest that we have to give up the classical logic of quantification.⁵⁰

As a further consequence of our analysis of the paradoxes of indirect discourse, we know that any approach that allows for a truth predicate satisfying the unrestricted principle (TB) will not provide an adequate treatment of the modalities and attitudes. For if (TB) is available, the paradoxes of indirect discourse can be formulated within this framework. Consequently, we can spell out the paradox within the quotation name approach, even though this approach allows for a limited amount of quantification only. We simply need to replace the premises by the appropriate variants of the quotation name language, i.e., (O1) is replaced by—reading \ddagger as a predicate

$$\ddagger \forall x(\ddagger x \rightarrow \neg Tx)\}$$

⁴⁹If we use, e.g., the arithmetical names of sentences or propositions and work with an arithmetic base theory at least as strong as \mathcal{Q} this condition is trivially met. If a distinct class of quotation names is postulated, as we have seen, this need not be the case.

⁵⁰Asher [9] provides an inductive theory of propositional quantification which is based on Kripke’s theory of truth and which leads to replacing the axiom of universal instantiation by the corresponding rule of inference. A less drastic move would be to opt for a free logic of propositional quantification, but it’s not clear whether this really amounts to a viable alternative. Asher suggests that such a proposal would run into serious trouble with respect to anaphora (cf. pp. 22–23).

and (O2) becomes

$$\forall x(\dot{x}x \rightarrow x = \int \forall x(\dot{x}x \rightarrow \neg Tx)\int).$$

Since we know by Theorem 2.17 that we can consistently maintain (TB) within the quotation name approach this tells us that the quotation name approach dismisses the Prior situation as an impossible one. Thus it seems that the quotation name approach is not only unsatisfactory for allowing to little quantification, it also fails from a more general perspective, as it yields intuitively incorrect results with respect to the modal notions and especially the propositional attitudes. Moreover, this observation seems to extend to any approach built on classical logic of quantification in which a truth predicate governed by the unrestricted principle (TB) is available. The moral of this observation seems to be that any approach to truth and the modalities, which is adequate from a natural language perspective, will require the objects to which truth and the modal notions are attributed to be very finely structured and moreover, their structure will need to be transparent within the approach.⁵¹

This affects all treatments of modality and truth, no matter whether we conceive of these notions as predicates of sentences or propositions. Nor will it help to conceive of the modalities as operators, if we allow ourselves to the resources of propositional quantification.

Not all modal notions are clearly in need of a fine grained individuation of propositions, or, more generally, the objects the modal notions are ascribed to. Accordingly, the champion of the operator view might be tempted to dismiss the paradoxes of indirect discourse by arguing that these are in fact paradoxes of direct discourse and reducible to the Liar paradox. The view would be that, e.g., ‘Onephrase asserts’ is in fact a predicate of sentences and the quantifier just a classical first-order quantifier quantifying into the argument position of a sentential truth predicate. Whether an inconsistency arises then depends entirely on the properties of the truth predicate and thus the paradox reduces to the liar paradox.

In the case of modal notions like alethic possibility it seems sufficient to conceive of propositions as sets of possible worlds. But, as it is well known, this individuation already runs into trouble when it comes to propositional attitudes like knowledge and belief, as it implies that logically equivalent sentences express one and the same proposition. Thus, for instance, assuming for sake of the argument the Goldbach conjecture to be true, this implies that if one knows that $2 + 2 = 4$, one equally knows the Goldbach conjecture. Even though the sets of possible worlds conception

⁵¹Compare the gloss Brandenburger and Keisler give of there “impossibility theorem”:

No belief model can be complete for a language that contains first-order logic. That is every belief model has holes expressible in first-order logic, where a hole is a statement that is possible but never assumed by the other player. (See Brandenburger and Keisler [19], p. 213)

A belief model in Brandenburger and Keisler’s sense is some possible world model.

of propositions has its proponents, we can't see how, given this individuation of propositions, propositions can do the job their advocates want them to do.

Other accounts of propositions—still coarse grained enough to avoid the paradoxes—might fare better in this respect, as they might not imply that logically equivalent sentences express the same proposition. Such modelings of propositions might be adequate with respect to the alethic modalities and some propositional attitudes such as belief and knowledge. But even granted the adequacy of these modelings of propositions with respect to a wide class of modal notions it is not clear how this could save the operator approach, or, more generally, the indirect discourse approach, from the paradoxes because it seems difficult to argue that all versions of the indirect discourse paradoxes reduce to direct discourse, i.e., liar like paradoxes. If one has bought into the operator view in the first place, why should 'thinks to himself' count as direct discourse while, presumably, 'believe' does not?⁵² It's not clear what a non question begging distinction between modal operators or modal predicates of coarse grained propositions, and modal predicates of finely structured entities—or, alternatively, direct and indirect discourse modalities—would look like in light of these remarks. Clearly, a criterion for such a distinction would need to tell us why amongst the modal notions we need to distinguish between operators and predicates whereas at the outset all these expressions seem to be expressions of the same category.

But even if one can argue for the above view it is questionable whether this would help with respect to the paradoxes of indirect discourse. Since ultimately it seems that an adequate treatment of modalities should provide a framework which allows for a treatment of several modalities and propositional attitudes simultaneously, and in which these modalities interact. But in this case we will have to make sense of sentences like

Everything Prior thinks to himself (to be true) Prior believes.

while it's not clear what we should take the logical form of this sentence to be. We can take "is believed by Prior" to either be a predicate of coarsely individuated propositions or a modal operator. In the first case, we have to deal with two sorted quantification and we have to explain over what entities the natural language quantifier ranges. The obvious answer is to introduce a predicate for the "expresses-relation" into the language, i.e., the predicate 'sentence x expresses the proposition y ' into the formal framework. Provided a two sorted language with two quantifiers \forall^S and \forall^P and variables x, x', x'', \dots and p, p', p'', \dots , a predicate $Expr(x, p)$ linking the two sorts of terms and the two modal predicates \ddagger and Bel where the former is

⁵²See Prior [124] and Thomason and Tucker [171] for further argument along these lines. Most persuasively, Prior argues that indirect discourse paradoxes can be construed for a wider class of attitudes as one might think. One example which Prior (cf. [124], p. 18) attributes to Geach is the attitude 'A schizophrenic fears that'. Assumption(i) would then become: A schizophrenic fears that everything a schizophrenic fears is not the case.

a predicate of sentences and the latter one of propositions,⁵³ we can formalize the sentence by

$$\forall^S x \forall^P p (\ddagger x \wedge Expr(x, p) \rightarrow Bel(p))$$

In the second case where we treat ‘Prior believes’ as a modal operator, the natural language quantifier ‘everything’ seems to quantify into term and sentence position simultaneously. The obvious way to achieve this is to introduce what Belnap [13] calls a subnector: a device which takes sentences as arguments to produce terms. Given such a device and a propositional quantifier \forall as well as a modal predicate \ddagger and a modal operator \Box we can formalize as follows

$$\forall p (\ddagger Q(p) \rightarrow \Box p).$$

But obviously, in both cases we reintroduce the paradoxes through the backdoor, if no restrictions on the expressibility predicate or the subnector are put forward.⁵⁴ And it seems that strategies, which place restrictions on the expressibility predicate or the subnector will, again, fail to be adequate with respect to our natural language intuitions. Therefore, the strategy of treating some modalities as applying to coarse grained objects, that is, as modalities of indirect discourse does not seem to amount to a viable proposal. Consequently, we need a fine grained individuation of the objects of the modal ascriptions, as in our framework we allow for a wider range of modalities simultaneously, which we treat in a homogeneous way, that is, by expressions of the same grammatical category.

In light of these remarks we conclude that an adequate treatment of quantifying into the argument position of modal notions, makes these notions vulnerable to paradox. Tentatively we suggest that the reason for this lies in the fact that by quantifying into the argument position of the modal notion a sentence or a proposition can refer back to itself and thus quantification introduces self-reference. If the structure of the sentence or the proposition is finely individuated, then this self-reference might lead to an infinite regress similarly to the one familiar from the liar paradox. Accordingly, any account of modalities hoping to be adequate with respect to natural language will either have to adopt weaker principles than those commonly adopted in modal logic, or put restrictions upon the classical logic of quantification. The former option applies to the predicate account whereas the latter is the road to take for the operator view.

2.3.3 Modalities, Reification and Self-reference

We have argued that once the means to quantify into the argument position of modal notions be it by appeal to propositional quantifiers in the operator case or standard

⁵³Note, that since by presupposition the propositions are individuated coarsely, the belief predicate can obey the standard laws of doxastic logic.

⁵⁴Already, Thomason [169] and Asher and Kamp [11] warn that paradox might be introduced via the expresses-relation, if no special care is taken.

first-order quantification in the predicate case are available, a certain amount of back or self-reference is introduced, which might, if further conditions are met, lead to paradox.

Independently of quantification the capacity of referring back to certain assertions—if not to the assertion under consideration itself—seems to be highly desirable, if one hopes to provide an adequate account of truth and the modalities from a natural language perspective. Natural language possesses devices such as demonstratives, anaphora and, more generally speaking, pronouns which are designed to fulfill this task. Moreover, these devices have the effect of reifying assertion, sentences or propositions. That is, these devices transform assertions, sentences or propositions into objects of discourse. Objects we can then speak about.

Independent evidence for this thesis resides in the fact that the paradoxes are not isolated phenomena of formal languages but can be formulated in natural language. Again, this conclusion seems to be at least *post* Kripke widely accepted in discussions focusing on notion of truth, but commonly neglected when it comes to the modalities. However, as the liar paradox can be reconstructed in natural language so can the modal paradoxes. This can be witnessed by the following reconstruction of the Paradox of the Knower:

Consider the sentence

I don't know this sentence

and call it KN. Now, let's assume that I know KN. Then by the factivity of knowledge, i.e., the fact that everything that is known is the case, I can infer KN. But KN says that I don't know *this sentence*. But *this sentence* just is KN and hence I don't know KN. We have derived a contradiction starting from the assumption that I know KN. Accordingly, it seems sound to conclude that I don't know KN and it even seems that I have just produced an impeccable proof to the effect that I don't know KN. But then, since I have proven that I don't know KN, I seem licensed to conclude that I know that I don't know KN. Thus I know the sentence that I don't know KN. But *the* sentence that I don't know KN is just KN itself and therefore I can conclude that I know KN and we have ended up in contradiction.⁵⁵

Upon reflection this natural language rendering of the knower paradox relies besides the factivity of knowledge on two features of natural language, which are independent of this particular example. First, as mentioned natural language allows for reifications of sentences. That is, we can transform sentences into objects of discourse. In the natural language paradox this is achieved by naming. We gave a sentence the name 'KN'. Whereas this might be bizarre in the present example note that giving names to sentences or other expressions is a common practice even though, most of the time they come as definite descriptions, witness 'the sentence written on the blackboard' or 'Bob's assumption'. This naming policy allows us to speak about the properties of sentences or propositions and to refer to these sentences via pronouns.

Moreover, this makes it possible to represent substitutions of expressions which appears to be the second crucial feature of natural language enabling us to formulate

⁵⁵Cf. Tymoczko [181] for reconstructions of the paradoxes along these lines.

the paradoxes, i.e., we repeatedly equated the demonstrative or descriptive phrase ‘this sentence’ with ‘KN’ and thereby concluded that we have derived a contradiction, but this conclusion is only warranted if we allow for substitutions of expressions.

Perlis [121, 122] is rather explicit on this behalf. He argues that in general substitutivity of expressions is representable in natural language.

A further feature of natural language, and one that effective self-reference appears to hinge upon, is that of substitutivity. By this I mean the ability to refer to the result of making alterations in a statement, such as “If you had said *John is here* instead of *Mr. Smith is here*, I would have understood who you meant.” (See Perlis [122], p. 184. The italics are due to Perlis)

The moral for a formal treatment of modalities which ultimately aims at taking natural language seriously is that it should provide the resources necessary to at least refer to and speak about sentences and, moreover, it should be capable of expressing substitution. The natural way to fulfill the former requirement is to have names of sentences in the language, which, additionally, would have the effect that first-order individual variables could be placeholders for sentences of the language in the evaluation of the formulas. The alternative would be to introduce propositional variables and dispense of names of sentences. But again, in both cases once we introduce substitution into the framework paradoxes will arise.

From a formal point of view substitution is a function from formulas to formulas which can be represented within the object language either by a sentence to sentence, i.e., a three place modal operator from triples of formulas to formulas, or by term to term devices. If we work in a language and theory in which a sufficient amount of syntax theory is available, a term to term device expressing the substitution function can in fact be defined. This is, of course, known since the work of Gödel and, as we have discussed, if the modal notion is conceived as a predicate and thus applicable to terms likewise, Montague showed that certain combinations of modal principles lead to paradox.⁵⁶

If substitution is represented by a sentence to sentence device, the situation parallels the case of quantification. That is, paradox will arise independently of the modal principles assumed. Let $Sub(\cdot, \cdot, \cdot)$ be a substitution operator where $Sub(\phi, \chi, \psi)$ is supposed to be the formula resulting from ϕ by substituting ψ for χ in ϕ . However, for ease of presentation we shall only consider a two-place p -substitution operator, that is substitutions where χ is the propositional variable p and we call the resulting operator $PSub$. The most straightforward way to capture the intended reading of $PSub$ is by postulating an unrestricted schemata of substitution⁵⁷:

$$(SB) \quad PSub(\phi, \psi) \leftrightarrow \phi(\psi/p).$$

Clearly, if we adopt this schemata we end up in inconsistency for by a simple application of (SB) we may derive

$$(NSB) \quad PSub(\neg PSub(p, p), \neg PSub(p, p)) \leftrightarrow \neg PSub(\neg PSub(p, p), \neg PSub(p, p))$$

⁵⁶Modulo the necessary qualification we have discussed in the present chapter.

⁵⁷Cf. Perlis [122] pp. 186/187.

Still, in this case the machinery has to be introduced, that is, substitution cannot be defined within the language as it is when represented by a term to term device assuming a sufficient amount of syntax theory.

The reason why substitution may introduce paradox is that it allows us to diagonalize formulas in which a variable, be it an individual or a propositional variable, occurs free. That is, let, e.g., $\phi(p)$ be a formula with a free propositional variable p . Then we may, using the substitution operator, always find a formula χ such that

$$\phi(\chi) \leftrightarrow \chi.$$

This follows directly from (SB) because as an application of this schemata we have

$$\phi(PSub(\phi(PSub(p, p)), \phi(PSub(p, p)))) \leftrightarrow PSub(\phi(PSub(p, p)), \phi(PSub(p, p))).$$

Now, if we remain in the operator setting and let $\phi(p)$ be the formula $\neg p$, we obtain the inconsistent (NSB). We cannot diagonalize this formula, if we return to the first-order setting where substitution is a term to term device. In a nutshell, the difference between the two forms of substitution arises because if substitution is expressed by a sentence to sentence device, we can assert the results of the substitutions unrestrictedly. Substitution behaves like a modal operator and this operator applied to formulas yields a new formula. If substitution is expressed by a function symbol, which operates on terms, we need to appeal to transition principles, principles, which allow us to go from the fact that a certain predicate, e.g., ‘is true’ or ‘is necessary’, applies to the name of the sentence which is the result of a substitution to an assertion of this very sentence itself. Indeed, it is exactly what the standard laws of modal logic seem to do, if we formulate them in the predicate setting, i.e., they allow us to move from merely talking about substitutions to actually asserting them. Viewed from this perspective the liar paradox and the modal paradoxes might indicate to which extent we may assert the result of substitutions within the framework and as such the paradoxes might point us toward the limitations of the framework adopted.⁵⁸

So far we have argued that any formal approach to truth and the modalities that aims to be adequate from a natural language perspective should have the resources of representing substitutions of expressions and thus resources to talk about these substitutions although this is not to be conflated with asserting or stating the result of a substitution. Now, as we have just discussed, within the first-order setting one role of predicates like ‘is true’ or ‘is necessary’ seems to be that they allow for the transition from merely talking about some substitution to its assertion. And this transition is mediated by the modal principles we adopt. But then the proper setting for developing a modal logic seems to be one where we can, at the very least, diagonalize with respect to the argument position of the modal notion at stake. This is most naturally done, for reasons we have just sketched, when we conceive of the modal notion as a predicate but can be done for the operator setting likewise. The latter has been explored by Smorynski [154, 155] under the label of Diagonalization Operator Logic (DOL).

⁵⁸See Yanofsky [181] and, following Yanofsky, Abramski et al. [1] for comments along these lines.

Smorynski introduces new fixed-point operators into the modal language, operators that applied to a formula $\phi(p)$ where the propositional variable p occurs only within the scope of the modal operator yield the diagonalization, i.e., the fixed-point $\delta_{\phi(p)}$ of this formula. Thus, in DOL we may prove

$$\phi(\delta_{\phi(p)}) \leftrightarrow \delta_{\phi(p)}$$

for every formula $\phi(p)$ where p occurs only in the scope of the modal operator.

We shall discuss Smorynski's DOL in some detail in the next chapter but the upshot is that it enables us to compare the effects of diagonalization, with respect to the modal operator and the modal predicate. And, as we shall see, it shows that predicate and operator are on a par once we can diagonalize the modal operator: the same modal principles lead to paradox.⁵⁹ Consequently, we better had a modal logic that does not lead to inconsistency in such a setting.

2.4 Conclusion

In sum, we have argued in this chapter that whether modalities are treated as operators or as predicates is orthogonal to the question of paradox. In this we concur with, amongst others, Perlis [121, 122], Grim [51], Belnap and Gupta [14], and Egré [38]. This conclusion is supported by our observation that

- we can save all of propositional modal operator logic, contra Montague, when we treat modalities as predicates. Moreover, the modal predicates under consideration are proper predicates, that is, no syntactic restrictions have been put forward with respect to these predicates;
- if we equip the modal framework with the means necessary to quantify into the argument position of the modality under consideration, the paradox will arise whether we treat the modality as an operator or as a predicate as long as the interpretation remains faithful to our natural language intuition;
- if the modal framework possesses the means necessary to speak about and to refer (back) to sentences, and can express substitutions, then the framework is vulnerable to paradox no matter how we conceive of the modality. We argued that

⁵⁹This claim is probably a bit too strong. However, all the standard inconsistency (this is meant to exclude ω -inconsistency) results which are derived via application of the the simple Diagonal lemma or simple diagonalization, can then be reproduced in the operator setting. Egré [38] pursues the parallel between the modal antinomies arising within the approach to modality and the inconsistency results obtained, if DOL is taken as the base modal logic and concludes:

Although the results given above [Inconsistency results for syntactical treatments of modality. J.S.] involve metalinguistic predicates and use the resources of first-order arithmetic, most of them are naturally seen as inconsistency results for corresponding propositional modal logics with fixed-point operators. (See Egré [38], p. 30).

from a natural language perspective an adequate account of modalities should be equipped with these expressive resources.

This shows that we should not shy away from syntactical or predicate treatments of modality because of the paradoxes but rather take the paradoxes as an indication that we are on the right track. But, of course, this is no argument that we should opt for an predicate approach to modality, as we could supplement the operator approach with the necessary expressive means.

However, while we have not provided an argument for a predicate as opposed to an operator approach to modality, we do think that conceiving of the modalities as predicates comes very natural once we are interested in an account of the modalities, which is expressively rich in the above sense. We may list the following reasons for our assessment:

- Quantification, substitution and thus self-reference with respect to the modal notion is most commonly and perhaps naturally expressed in the setting of first-order logic.
- Assuming that truth is formalized as a predicate, the predicate approach to modalities would guarantee that the concept of truth and the modalities are treated on a par, which seems to be a rather basic linguistic intuition.

There are further but rather pragmatical reasons why it might be preferable to work with a modal predicate instead of a modal operator, if one is interested in an expressively rich framework. For on the one hand, due to the rich and in-depth literature on the truth predicate and the liar paradox, we already possess a firm starting point for developing such an account and a large body of mathematical knowledge to rely on. On the other hand, we face the completely opposite situation in the operator case for practically no work on modal operator languages and logic, which are expressively rich in the relevant sense, has been conducted. We would thus need to start our research from scratch.

Of course, we are not always in need of the full expressive power sketched above. Often standard modal operator logic seems to do the job quite fine and it would be foolish to attribute a mistaken conception of modality to all the highly successful research and applications of standard modal operator logic. From this consideration emerges an immediate consequence for any predicate approach to modalities, namely, that it should be at least consistent with the standard operator accounts of modality. This leads us to the general question what an adequate account of modalities conceived as predicates could look like.



<http://www.springer.com/978-3-319-22556-2>

Toward Predicate Approaches to Modality

Stern, J.

2016, VIII, 190 p., Hardcover

ISBN: 978-3-319-22556-2