The Liar Paradox and Modalities

Johannes Stern
Department of Philosophy
University of Bristol
johannes.stern@bristol.ac.uk

May 20, 2023

1 Introduction

The Liar paradox shows that some of our basic intuitions regarding truth seem to lead us straight into contradiction. It seems that we can prove that the Liar sentence—a sentence that says of itself that it is not true—is true, but also that it is false.[1] For many theorists—pace advocates of paraconsistency—this seems to be an unacceptable conclusion. What is perhaps less well known is that similar kind of paradoxes affect a variety of modal notions and propositional attitudes. Let ml be the sentence in the next line, i.e., the sentence

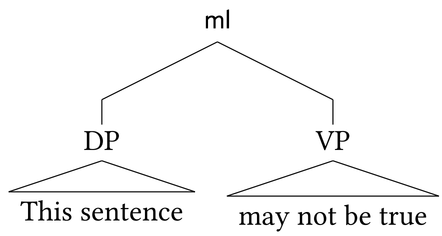
This sentence may not be true

Now assume ml is false, i.e., $\neg\mathsf{ml}$. This tells us that

It is not the case that ml may not be true.

Hence, assuming the modal square of oppositions, ml must be true and we can infer ml. We derived a contradiction from the assumption $\neg\mathsf{ml}$ and have thus proved ml by reductio. But since we have proved ml we know that it must be true, yet ml says that it may not be true. By modal square of oppositions the latter implies that it is not the case that ml must be true. We have arrived at a contradiction.

The syntax of the sentence ml is typically parsed as



On the assumption that VPs are formalized using primitive or complex predicates the sentence

[1]Some theorists will hold that to say that a sentence is true or false amounts to a category mistake and is hence unintelligible. The idea is that what can be true or false is whatever is expressed by a given sentence in a given context, e.g., a proposition. But note that semanticists do apply the truth predicate to sentences of the language and thus it would seem that the truth predicate can be meaningfully applied to sentences.

δ such that

(†) $$\delta \doteq \mathrm{P}\ulcorner\neg\delta\urcorner$$

seems a natural formalization of ml where $\ulcorner\delta\urcorner$ is a name of the sentence δ.[2] 'P' is a one-place predicate whose intended reading is 'may be true' or 'is possible'. Given this formalization one can rigorously derive a contradiction from the two following schematic assumptions

(T_N) $$\mathrm{N}\ulcorner\varphi\urcorner \longrightarrow \varphi;$$

(Nec_N) $$\frac{\varphi}{\mathrm{N}\ulcorner\varphi\urcorner}$$

combined with the definition of the predicate 'N' (read 'must be true' or 'is necessary')

(‡) $$\mathrm{N}\ulcorner\varphi\urcorner :\leftrightarrow \neg\mathrm{P}\ulcorner\neg\varphi\urcorner.$$

To this effect assume δ is false, i.e., $\neg\delta$.[3] Then

1. $\neg\delta$;
2. $\neg\mathrm{P}\ulcorner\neg\delta\urcorner$, 1, definition of δ;
3. $\mathrm{N}\ulcorner\delta\urcorner$, 2, (‡);
4. δ, 3, (T_N);
5. $\bot$, 1,4.

We have derived a contradiction from the assumption $\neg\delta$ and by reductio we conclude δ. Since we have proved δ we can use the rule (Nec_N) and derive $\mathrm{N}\ulcorner\delta\urcorner$. Then

1. $\mathrm{N}\ulcorner\delta\urcorner$
2. $\neg\mathrm{P}\ulcorner\neg\delta\urcorner$, 1, (‡);
3. $\neg\delta$, 2, definition of δ.

We have arrived at a contradiction. It seems that if in our formal framework we allow for formal counterparts of sentences like ml contradiction ensues, if modal principles such as (T_N) and (Nec_N) are assumed. As the informal presentation of the paradox should hopefully make clear, these modal principles seem to be as constitutive for our understand of alethic modalities, knowledge and, perhaps, some factive propositional attitudes as the T-scheme is for the notion of truth. The interesting formal observation to these modal paradoxes is that the full

[2] Throughout this chapter we understand $\ulcorner\,\urcorner$ as a name-forming operation of a language $\mathcal{L}$, i.e., $\ulcorner\,\urcorner : \mathsf{Sent}_\mathcal{L} \longrightarrow \mathsf{Term}_\mathcal{L}$. If we are working in some arithmetical language we conceive of $\ulcorner\varphi\urcorner$ as the numeral of the Gödel number of φ denoted by $\#\varphi$. By $\doteq$ we denote that 'δ' are '$\mathrm{P}\ulcorner\neg\delta\urcorner$' are the same expressions. If we are working in an arithmetical language this requires the language to contain specific function symbols.

[3] Throughout this chapter we will understand rules such as (Nec_N) as rules of proof and *not* as introduction rules of a natural deduction system.

strength of T-scheme is not needed in the derivation of the Liar-like paradoxes. In the modern era this was arguably first noticed by Montague, Myhill and Kaplan (Kaplan and Montague, 1960; Myhill, 1960; Montague, 1963) who showed that various combination of modal principles, derivable by way of the T-scheme, lead to inconsistency if self-referential sentences or propositions are available in the formal framework. The most prominent result establishes the joint inconsistency of the principles (T_{N}) and ($\mathsf{Nec}_{\mathrm{N}}$) we have just presented, but there are further and arguably more surprising inconsistency results which even affect non-factive notions such as belief (see, e.g., Thomason, 1980; Koons, 1992). As a whole the family of modal, i.e. Liar-like, paradoxes, provides us with an interesting insight into the prospects and limitations of formal treatments of modalities and propositional attitudes. In Part I of this chapter we attempt to systematize the various paradoxes and inconsistency results to provide an overview of the options and limitations the modal paradoxes pose for formal treatments of modality.

In light of our example some readers will protest against the idea of modal paradox. They might hold that the paradoxicality of the sentence ml is not to be blamed on the modal notion but on the notion of truth, i.e., the paradoxicality of ml is down to the truth predicate that occurs in the sentence. Indeed, the understanding the possibility predicate in terms of 'may be true' suggests that the predicate should be understand as a modified truth predicate and that the paradox is not due to the modality modifying the truth predicate, but the truth predicate itself. We pick up on this idea in Part II of the chapter where we turn to strategies in way of satisfactory formal accounts of modal notions and propositional attitudes. The guiding idea behind the various strategies discussed is to characterize modal notions and propositional attitudes via their interaction with other such notions, and their interaction with provability and truth. Focusing on the interaction with the notion of truth leads to a particularly attractive picture according to which the modal paradoxes "reduce" to the paradoxes of truth (the Liar paradox). On this view, once we have devised a consistent (non-trivial) theory of truth an attractive formal account of modality and propositional attitudes can be given.[4] The two parts of the chapter can be read independently of each other, although a reader of Part II may find the occasional reference to the paradoxes discussed in Part I.

Before we start it may be worth addressing a general idea that has been voiced implicitly and explicitly in the relevant literature: some theorists have taken the modal paradoxes to show that modal notions and propositional attitudes ought to be formalized by sentential operators as in modal operator logic, and semantically conceived of as quantifiers ranging over situations (possible worlds).[5] If the modal predicate P (N) is replaced by a modal operator $\Diamond$ ($\Box$), then the name of a sentence cannot occupy the argument position of the modal notion, which, at least prima facie, seems to block the derivation of the paradoxical conclusion. Constructing a self-referential sentence like ml via naming is no longer possible, if modalities are conceived as operators: if P were replaced by $\Diamond$ in (†), we would obtain the expression $\Diamond\ulcorner\delta\urcorner$ which is not a well-formed formula. On this view, the paradoxes are avoided by formalizing modalities,

[4] Unsurprisingly, if some non-classical logic is adopted in reaction to the Liar paradox, the same logic can be used to spell out a modal logic in which the naive intuitive principles are preserved to the extent the non-classical logic enables maintaining the T-scheme. In this chapter we do not discuss non-classical approaches to modal paradox, as we feel there is nothing interesting to say that is not already covered by the discussion of non-classical approaches to the Liar paradox. This said, much of discussion in Section 5 generalizes to the non-classical case.

[5] See Slater (1995) for a pronounced statement of this conclusion.

knowledge, propositional attitudes as sentential operators. This kind of reaction to the modal paradoxes is problematic to say the least. There may be good arguments to conceive of modal notions and, perhaps, propositional attitudes as sentential operators or situation quantifiers, but the paradoxes are no such argument. For one, such a reaction would be totally asymmetric to the reaction in case of the Liar paradox: the research community has not taken the Liar paradox as an argument in favor of a truth operator. For another, the natural language derivation of the modal paradox suggests that in any suitably expressive framework in which rich fragments of natural language can be represented sentences like ml can be constructed. This means that if one were to formalize modalities and propositional attitudes as sentential operators, then one needs to find alternative means for constructing sentences like ml. In conclusion, the modal paradoxes should not be dismissed out of hand, but rather seen as a challenge any interesting account of modality and the propositional attitudes needs to address.[6]

Part I
The Dark Side: Modal paradox

Modal paradox seems to show that some of our fundamental intuitions regarding our understanding of modal notions are inconsistent. It shows that if certain unrestricted schemata characterizing truth, modal notions, or propositional attitudes are assumed, paradox will ensue. Whereas the Liar paradox shows that the unrestricted T-scheme,

$$\mathrm{T}\ulcorner\varphi\urcorner \leftrightarrow \varphi,$$

leads to paradox, the Modal liar establishes that the paradoxical conclusion still follows if the right-to-left direction of the T-scheme is replaced by a rule of proof. A natural question arises in this context: which combinations of modal principles are consistent and which are inconsistent?

Given the plethora of different possible schemata a comprehensive and complete answer to this question is of course impossible, yet one can still aspire to provide some systematization.[7] To this effect, it seems promising to take the lattice of modal operator logics as a starting point and to investigate which modal logics will be consistent/inconsistent, if self-referential sentences like ml are introduced to this setting. Of course, in standard modal operator logic such self-referential sentences will typically not be available. In the derivation of the paradox in the Introduction to this chapter we conceived of modality as a predicate and stipulated the existence of such a sentence by fiat. A more systematic approach is to assume some syntax theory in the background that provides a sufficiently rich naming system, or a theory of finely

[6] See Stern (2014c, 2016) for discussion of the moral of the modal paradoxes and of related inconsistency results. Of course, in many cases it seems appropriate to work in expressively weak systems and to disregard complications arising because of semantic or intentional paradox.

[7] Systematization and discussion of various paradoxes may be found in, e.g.,Friedman and Sheard (1987); Egré (2005); Schwarz (2013); Stern (2016). In this chapter we set discussion of the probalistic liar and its impact on probabilism aside, but see Caie (2013); Campbell-Moore (2015a,b, 2016) for discussion.

grained structured propositions.[8] For example, this could be an arithmetical theory like PA, some weak set theory or a syntax theory proper such as the one developed in Halbach and Leigh (2022). If such a theory Σ is assumed in the background one can show that for every formula $\varphi(v)$ with exactly one free variable, there exists a sentence γ such that

$$\Sigma \vdash \varphi(\ulcorner \gamma \urcorner) \leftrightarrow \gamma.$$

This is Gödel's well-known diagonal lemma. The diagonal lemma can be used to construct the sentence δ we used in deriving the modal paradox in the Introduction to this chapter in a mathematically precise way.[9]

If we wish to work with the lattice of modal operator logics this systematic approach for constructing self-referential sentences such as ml is not available to us, as the diagonal lemma enables us to diagonalize with respect to the term position of a formula only, i.e., the diagonal lemma is not applicable to a formula $\neg\Box p$ where $\Box$ is a sentential operator. The syntax of the modal operator language is specified as follows:

$$\varphi ::= p_i \,|\, \neg\varphi \,|\, \Box\varphi \,|\, \varphi \wedge \varphi \,|\, \varphi \vee \varphi$$

where p_i with $i \in \omega$ is a propositional variable. We conceive of all other boolean connectives as defined and as mere notational abbreviations of their definiens. Similarly, we think of $\Diamond\varphi$ as short for $\neg\Box\neg\varphi$. To construct a diagonal sentence for the formula $\neg\Box p$ further machinery needs to be added to the language and logic of modal operators. There are various ways how this can be done, but, perhaps, the easiest way is to use Smorynski's (Smorynski, 1985, 2004) Diagonal Modal Logic (DML). DML extends modal operator language by fixed-point constants δ_φ for all formulas $\varphi(p)$ such that the propositional variable p occurs only in the scope of the modal operator in φ. DML extends modal operator logic by stipulating for each such formula $\varphi(p)$ the so-called fixed-point axiom

$$\varphi(\delta_\varphi) \leftrightarrow \delta_\varphi.$$

These fixed-point axioms mimic instances of the diagonal lemma in modal operator logic. For example, δ as introduced in (†) is obtained via the fixed-point axiom

$$(\mathrm{FP}_{\Diamond\neg p}) \qquad \Diamond\neg\delta_{\Diamond\neg p} \leftrightarrow \delta_{\Diamond\neg p}.$$

It is then straightforward to show that the modal operator versions of (Nec), (T) are jointly inconsistent with $(\mathrm{FP}_{\Diamond\neg p})$ and the definition of $\Box$-operator as $\neg\Diamond\neg$. Throughout the chapter we

[8]From a philosophical perspective it might be preferable to think of modal and attitudinal predicates as applying to propositions rather than sentences—at least if these predicates are understood as primitives as opposed to complex predicates. Prima facie, conceiving of 'belief' as a primitive sentential predicate seems to amount to analysing belief as a relation between agents and syntactic objects, which, at least intuitively, seems wrong.

[9]Notice however there is a further assumption that enters the derivation of the paradox and which often goes by unnoticed: the naming system that is used to state the diagonal lemma and to find self-referential sentences needs to also be employed in the formulation of the modal (truth) schemata. If, in contrast, a naming system that is expressively impoverished is used in formulating the modal principles, paradox may be avoided even if the diagonal lemma can be proved for some richer naming system which is available. For discussion, further details and the development of modal logics in such a setting see, e.g., Niemi (1972); Gupta (1982); Visser (1989); Asher and Kamp (1989); Schweizer (1992); Stern (2014c, 2016).

frequently drop the subscript of a fixed-point constant if no confusion can arise. Moreover, if L is a modal operator logic in the language without fixed-point constant, we call the extension of L by the fixed-point axioms the *diagonal extension* of L and denote it by L^F.[10]

2 Paradox and Normal Modal Logics

Normal modal logics are logics that extend classical propositional logic and are closed under the two modal principles[11]

(K) $$\Box(\varphi \to \psi) \to (\Box\varphi \to \Box\psi)$$

(Nec) $$\frac{\varphi}{\Box\varphi}$$

The minimal normal modal logic is called K. We can obtain different normal modal logics by adding further axioms to K. The modal axioms we shall appeal to are introduced in Figure 1. Note that for an axiom X we denote the converse direction of the axioms by X_c.[12]

D	$\Box\neg\varphi \to \neg\Box\varphi$
T	$\Box\varphi \to \varphi$
4	$\Box\varphi \to \Box\Box\varphi$
F	$\Diamond\varphi \to \Diamond\Box\varphi$
B	$\varphi \to \Box\Diamond\varphi$
E	$\Diamond\varphi \to \Box\Diamond\varphi$

Figure 1: Axioms of modal operator logic

Due to an algebraic argument by Makinson (1971) we know that normal modal logics fall into two (non-disjoint) camps: those that are *sublogics* of the so-called identity logic, that is, the logic axiomatized by the modal schema

($\mathsf{TB}_\Box$) $$\Box\varphi \leftrightarrow \varphi;$$

and those that are sublogics of the so-called unit logic axiomatized by the schema

(U) $$\Box\varphi.$$

For obvious reasons we will call the sublogics of the identity logic truth-like modal logics ($\mathbb{Tr}$) and, less obviously, we call sublogics of the unit logic provability-like modal logics ($\mathbb{Pr}$).

Proposition 1. *Let L be a normal modal logic. Then,*

(i) The modal logic K *is in* $\mathbb{Tr} \cap \mathbb{Pr}$*;*

[10] For a more in depth introduction to modal operator logic and diagonal modal logic we refer the reader to Smorynski (1985, 2004); Stern (2016).

[11] Throughout this chapter we assume logics to be presented in some (standard) Hilbert-style axiomatic calculus.

[12] For a general overview and introduction to normal modal logic see Blackburn et al. (2001).

(ii) $\neg\Box\bot \in L$ *iff* $L \in \mathbb{Tr} - \mathbb{Pr}$;[13]

(iii) *if* $\Box(\Box\varphi \rightarrow \varphi) \rightarrow \Box\varphi \in L$, *then* $L \in \mathbb{Pr} - \mathbb{Tr}$;

(iv) *if* $L \in \mathbb{Pr}$, *then* L^F *is consistent.*

Proof. (i) is immediate as Tr and Pr exhaust the family of *normal* modal logic and K is compatible with either interpretation of $\Box$. (ii) $L \notin \mathbb{Pr}$ there must be some $\chi \in \mathcal{L}$ such that $\Box\chi \vdash_L \bot$. By the deduction theorem we obtain $\vdash_L \neg\Box\chi$. Since for all normal modal logics $\neg\Box\chi \rightarrow \neg\Box\bot \in L$, it follows that $\neg\Box\bot \in L$. For (iii) assume $L \in \mathbb{Tr}$. Then $(\mathsf{TB}_\Box) \vdash \Box(\Box\varphi \rightarrow \varphi) \rightarrow \Box\varphi$ and in particular $(\mathsf{TB}_\Box) \vdash \Box(\Box\bot \rightarrow \bot) \rightarrow \Box\bot$. But $(\mathsf{TB}_\Box)$ proves $\Box(\Box\bot \rightarrow \bot)$, as $\mathsf{KD} \subseteq (\mathsf{TB}_\Box)$. We conclude $\Box\bot$ which contradicts KD.[14] (iv) see (Stern, 2016, Corollary 3.33). ☐

Proposition 1 informs us that the two families of modal logics are incomparable via the subset relation and that paradoxes arise only within the family of truth-like logics, which, however, is the family of logics of basically all interesting alethic modalities and propositional attitudes. The good news is that there are a couple of modal logics extending KD whose diagonal extension is consistent:

Proposition 2. *Let L be a modal logic with $L \subseteq S$ for $S \in \{\mathsf{KDD_c}, \mathsf{KD4_c}, \mathsf{KDF_c}\}$. Then L^F is consistent.*

Proof. The consistency of these diagonal extensions can be established by interpreting the modal operator logic in suitable first-order theories formulated over a suitable theory of arithmetics such as PA. The key idea is to understand (translate) the modal operator as a suitable sentential predicate, e.g., a truth predicate while keeping the interpretation of the boolean connectives fixed and assigning arbitrary sentences of the target language to the propositional constants. The fixed-point constants will be interpreted as suitable fixed points obtained via diagonalizations, i.e., the diagonal lemma. See (Stern, 2016, Ch. 3) for further details on the interpretation. The modal operator of $\mathsf{KDD_c}$ can be interpreted by the truth predicate of the theory FS (Friedman and Sheard, 1987, System D) and the modal operator of $\mathsf{KD4_c}$ by the truth predicate of System E of Friedman and Sheard (1987). The consistency of $\mathsf{KDF_c}$ can be established by interpreting the modal operator by a suitable Rosser provability predicate such as the one introduced in (Kurahashi, 2020, §4). ☐

Interestingly, for the logic $\mathsf{KDD_c}$ we obtain an even stronger result for it is possible to show that $\mathsf{KDD_c}$ is as strong as it is consistently possible: the diagonal extension of any logic properly extending $\mathsf{KDD_c}$ will be inconsistent.[15] Are there further such maximal normal operator logics?

These positive results contrast with key limitative results which suggest that attractive logics for alethic modalities and most propositional attitudes are out of reach, if fixed points for modal formulas are available.

[13] In normal modal logics $\neg\Box\bot$ is equivalent to the principle D.

[14] This is the modal version of Gödel's second incompleteness theorem (cf., e.g., Smorynski, 1985; Boolos, 1993, for discussion).

[15] See Czarnecki and Zdanowski (2019) for the original result and Nicolai and Stern (2021) for discussion of this fact. Nicolai and Stern also determine two non-normal maximal modal logics.

Proposition 3. *Let L, S be normal modal logics. Then*

(i) if $\mathsf{KT} \subseteq L$*, then* L^F *is inconsistent;*

(ii) if $\mathsf{KD4} \subseteq L$*, then* L^F *is inconsistent.*

(iii) if $S \in \{\mathsf{KDB}, \mathsf{KDE}, \mathsf{KDE_c}\}$ *and* $S \subseteq L$*, then* L^F *is inconsistent.*

Proof. (i) follows from the argument we have given in the Introduction to this chapter. For (ii) we observe that for L with $\mathsf{K4} \subseteq L$, $L^F \vdash \Box(\Box\varphi \rightarrow \varphi) \rightarrow \Box\varphi$. This follows by translating the proof of the formalized Löb theorem in to the modal operator framework (cf., e.g., Smorynski, 1985). The contradiction then follows by the reasoning of item (iii) in the proof of Proposition 1. The inconsistencies stated in (iii) can be derived by choosing suitable fixed points. For detailed proofs we refer the reader to (Ch. 3 Stern, 2016). □

The first item, i.e. item (i) of Proposition 3, is MONTAGUE'S THEOREM which we have already encountered, the second is a modal version of Gödel's second incompleteness theorem. One moral of the proposition is that most systems of modal logics that find application in philosophical arguments and discussions, e.g. $\mathsf{S4}$ and $\mathsf{S5}$, are inconsistent once fixed points of modal formulas are introduced to the system.

2.1 ω-Inconsistency and a Paradox of Common Belief

Unfortunately these limitative result are not the end of the story. If modal notions are treated as predicates, then we are working in a first-order setting over some reasonable syntax theory. In such a setting we have greater expressive strength than in modal operator languages. In particular we can quantify over argument position of the modality at stake and, to be sure, this seems to be highly desirable feature of the first-order setting and required to formalize philosophical thesis such as *knowledge is justified true belief* or *there are necessary a posteriori truths.* Yet, such quantification enables us to talk about all finite iterations of a modal predicate in front of some sentences, i.e., we can define an ω-modal predicate: in suitable syntax theories all primitive recursive functions can be represented and for ease of presentation we can even avail ourselves to function symbols of the language representing these functions. Let $f^\bullet$ be a function symbol representing the primitive recursive function f with

$$f(n, a) = \begin{cases} a, & \text{if } n = 0 \\ \#\mathrm{N}f^\bullet(\overline{m}, \overline{a}), & \text{if } n = m + 1. \end{cases}$$

Then the formula $\forall x(\mathsf{numb}(x) \rightarrow \mathrm{N}f^\bullet(x, \ulcorner\varphi\urcorner)$ intuitively asserts the infinite conjunction of all formulas

$$\underbrace{\mathrm{N}\ulcorner\mathrm{N}\ulcorner\ldots\mathrm{N}\ulcorner\varphi\urcorner\cdots\urcorner\urcorner}_{n+1\text{-times}}$$

for $n \in \omega$. Famously, McGee (1985) used this trick to show that the modal principles of the logic KD are ω-inconsistent if the principle

(#BF) $$\forall x(\mathsf{numb}(x) \rightarrow \mathrm{N}\ulcorner\varphi(\dot{x})\urcorner) \rightarrow \mathrm{N}\ulcorner\forall v(\mathsf{numb}(\mathsf{v}) \rightarrow \varphi(v))\urcorner$$

is assumed.[16] The intended reading of the predicate 'numb' is 'is a natural number'. Such a predicate can be defined in every adequate syntax theory. In combination with an ω-modal predicate, say 'is ω-necessary', (#BF) asserts that if some sentence is ω-necessary, then it is necessarily so.

Proposition 4 (McGee)**.** *Let* Σ *be a theory extending Robinson arithmetic that proves* $\mathsf{K_N}$, $\mathsf{D_N}$, #BF *and that is closed under* $\mathsf{Nec_N}$*. Then* Σ *is* ω*-inconsistent.*

For the proof of the Proposition we refer the reader to McGee (1985); Halbach (2014). We shall give an alternative presentation of McGee's ω-inconsistency result in due course but before it might be good to pause and reflect why McGee's result is problematic—after all (#BF) is not necessarily a plausible assumption, especially if one thinks of the modal predicate as an attitudinal predicate such as a belief or a knowledge-predicate. (#BF) is a form of formalized ω-rule which says that if all number instances $\varphi(\overline{n})$ are believed (known), so is its universal closure. It is simply too strong an assumption to take an agents beliefs or knowledge to be closed under the ω-rule.[17] Still, it is one thing to refrain from closing an agent's beliefs or knowledge under the ω-rule but something very different to commit one to the view that they cannot be consistently closed under the ω-rule—and these remarks apply even more strongly to alethic modalities.[18] Yet, McGee's result shows the latter. It show that we cannot coherently conceive of ideal situations in which an agents belief are closed under the ω-rule, that is, unless we are happy to accept an ω-inconsistent system of belief.

There are thus good reasons why an ω-inconsistent system of belief is to be avoided. Indeed, switching back to sentential operators McGee's ω-inconsistency can be cast as paradox of common belief: if KD is assumed as an agent's logic of belief, it is impossible to introduce a coherent notion of common belief along the usual lines.[19] The common belief of a (finite) group of agents G is typically thought of as the infinite conjunction of all the groups shared beliefs (cf. below) and common belief is thus a kind of ω-modality definable by McGee's trick. There are then two ways one can present McGee's ω-inconsistency result as a paradox of common belief. One can either (i) extend the syntax of the language to allow for infinite conjunctions and *define* the common belief operator or (ii) introduce a primitive common belief operator to the language together with rules and principles that reflect our intuitive understanding of common belief as an ω-modality.

Let's make these remarks precise and introduce the notion of common belief. To this effect we consider not one but a family of modal operators $\Box_j$ for each agent j of a finite group G. $\Box_j$ is to be read as 'agent j believes'. We then define an shared belief operator, E_G, relative to a group G, which says that 'every agent of the group believes'

$$\mathsf{E}_G(\varphi) :\leftrightarrow \bigwedge_{j\in G} \Box_j\varphi.$$

[16]A theory Σ is ω-inconsistent, if $\Sigma \vdash \varphi(\overline{n})$ for all $n \in \omega$ and $\Sigma \vdash \neg\forall x(\mathsf{numb}(x) \to \varphi)$. Furthermore, $\ulcorner\varphi(\dot{x})\urcorner$ denotes the (code of the) formula resulting from substituting the free variable in φ by the name of the value of x.

[17]The lottery paradox already shows that there are problems, if an agent's belief or knowledge is closed under finite conjunction. Closing them under infinite conjunction seems even more problematic.

[18]ω-inconsistency implies that there must be more numbers than the natural numbers, i.e., that there must be non-standard numbers that are not named by a numeral.

[19]For an introduction to common belief and its logic we refer the reader to Bonnano (1996); Fagin et al. (1995).

With this definition in place we can define iterative application of the shared belief operator as follows:

$$\mathsf{E}^0_G(\varphi) = \mathsf{E}_G(\varphi), \qquad \mathsf{E}^{n+1}_G(\varphi) = \mathsf{E}_G(\mathsf{E}^n_G(\varphi)).$$

Common belief is shared belief of all finite levels, i.e., that everyone believes that everyone believes that everyone believes… As mentioned, if our language has the expressive resources to formulate infinite conjunctions, then we can explicitly define the group's common belief operator C_G:

$$\mathsf{C}_G\varphi \;:\leftrightarrow \bigwedge_{n\in\omega} \mathsf{E}^n_G(\varphi).$$[20]

Notice that on this "definition" of common belief the implication

$$(\star) \qquad \mathsf{C}_G\varphi \longrightarrow \Box_j\mathsf{C}_G\varphi$$

can be proved for all $j \in G$. $(\star)$ is the operator counterpart to #BF: it says that if something is a common belief (ω-believed) then it is believed by an agent that it is a common belief. If common belief is conceived as primitive operator, $(\star)$ is a theorem of the logic of common belief.[21] With this in place we can derive McGee's ω-inconsistency result in guise of a paradox of common belief:

(1) $\delta \leftrightarrow \neg\mathsf{C}_G\delta$	fixed-point axiom
(2) $\Box_j\delta \leftrightarrow \Box_j\neg\mathsf{C}_G\delta$	1, (Nec), (K)
(3) $\Box_j\delta \longrightarrow \neg\Box_j\mathsf{C}_G\delta$	2, (D)
(4) $\mathsf{C}_G\delta \longrightarrow \Box_j\mathsf{C}_G\delta$	$(\star)$
(5) $\Box_j\delta \longrightarrow \neg\mathsf{C}_G\delta$	3, 4
(6) $\Box_j\delta \longrightarrow \delta$	1, 5
(7) $\neg\delta \longrightarrow \mathsf{C}_G\delta$	1
(8) $\neg\delta \longrightarrow \Box_j\delta$	7, $(\star)$
(9) δ	6, 8

[20] If G consists of only one agent j then $E_G\varphi$ is just $\Box_j\varphi$ and $\mathsf{E}^{n+1}_G(\varphi)$ is just $\Box^n_j\varphi$ and $\mathsf{C}_G\varphi$ amounts to the infinite conjunction of all $\Box^n_j\varphi$ for $n \in \omega$.

[21] There a number of different but equivalent axiomatizations. Here we follow Bonnano (1996) who assumes the axioms:

($\mathrm{S_i}$)	$\mathsf{C}_G\varphi \longrightarrow \Box_i\varphi$
($\mathrm{P_i}$)	$\mathsf{C}_G\varphi \longrightarrow \Box_i\mathsf{C}_G\varphi$
(L)	$\mathsf{C}_G(\varphi \longrightarrow \Box_1\varphi \wedge \ldots \wedge \Box_n\varphi) \longrightarrow (\Box_1 \wedge \ldots \wedge \Box_n\varphi \longrightarrow \mathsf{C}_G\varphi)$

for $i \in \{1, \ldots, n\} = G$, and the rule $\mathrm{RN}_{\mathsf{C}_G}$:

$$\frac{\varphi}{\mathsf{C}_G\varphi}$$

Bonnano's axiom ($\mathrm{S_i}$) is just our principle $(\star)$.

From Line 9 above we can infer $\mathsf{E}_G(\delta), \mathsf{E}_G(\mathsf{E}_G(\delta)), \dots$ In short, we can derive $\mathsf{E}^n_G\delta$ for every $n \in \omega$. If we avail ourselves to the infinitary rule

$$(\textstyle\bigwedge\text{-Intro}) \qquad \frac{\varphi_1, \varphi_2, \varphi_3, \dots}{\bigwedge_{i \in \omega} \varphi_i}$$

we can derive the definiens of $\mathsf{C}_G\delta$, if the latter is conceived of as an infinite conjunction. We then derive $\neg\delta$ by Line 1 contradicting Line 9. If the standard logic of common belief is assumed resorting to an infinitary rule is not necessary, as the logic of common belief is closed under the rule of necessitation for C_G (cf. Footnote 21). Then $\mathsf{C}_G\delta$ follows directly from Line 9. It seems that no coherent notion of common belief is available if the agent's logic of belief extends the modal logic KD. Ultimately, the result suggests that all truth-like modal logics need to be ruled out if we work in expressively rich frameworks in which fixed points of modal formulas can be constructed.[22]

This brings us to the end of our discussion of paradoxes and normal modal logic. The limitative results presented so far clearly indicate that satisfactory formal accounts of modalities, knowledge, and propositional attitudes will not be obtained by simply putting forward the principles of modal operator logics that are commonly assumed for the respective notions, as those will lead to inconsistency. Rather a new strategy is called for. Two immediate options come to mind.[23]

The first option is to characterize various modal notions by the way they interact amongst each other, and the notions of truth and provability. Ultimately, will take up this idea in Part II of the chapter. However, this path needs to be treaded with care, as new and surprising paradoxes may arise due to the interaction of different modal notions. As a matter of fact, the paradox of common belief we just discussed is one such paradox, at least if common belief is conceived of as a primitive.[24] The second option is, perhaps, more immediate and seeks to characterize the various notions using systems of non-normal logic. The idea is that the closure conditions imposed by (Nec) and (K) are simply too strong.

3 Paradox, Closure and Non-normal Modal Logics

The axiom (K) and the rule (Nec) guarantee that the *inner logic* of the modal notions, that is, the set $\mathsf{IN}_L := \{\varphi \mid L \vdash \Box\varphi\}$ is closed under modus ponens and theoremhood with respect to the given logic L. This means that all theorems of L are in IN_L and

$$\text{if } \varphi \rightarrow \psi \in \mathsf{IN}_L \text{ and } \varphi \in \mathsf{IN}_L, \text{ then } \psi \in \mathsf{IN}_L.$$

[22] In systematizing the various inconsistency results we assumed a syntactic, proof-theoretic perspective but, as Halbach et al. (2003) show, there is also a semantic side to these results: the inconsistency results can be naturally seen as undefinability theorems relative to possible world frames, e.g., Montague's theorem tells us that we cannot find a coherent interpretation/valuation on reflexive possible world frames.

[23] Of course, one option would also be to move to non-classical logics. However, as mentioned in the Introduction we will not discuss non-classical approaches in this Chapter. Let us point out however that, in general, it does not suffice to move to intuitionistic logic, as most inconsistencies can be reproduced in the intuitionistic setting (cf., e.g., Germano, 1970; Leigh and Rathjen, 2012).

[24] Further paradoxes of interaction are presented, e.g., by Halbach (2006, 2008) and Horsten and Leitgeb (2001). Whether all these paradoxes can be deemed novel and unexpected is, of course, open to debate (for discussion and some sharp results to this effect see, e.g., Stern and Fischer, 2015; Stern, 2016).

In the previous section we have seen that if such strong closure conditions are maintained we cannot find modal logics that rule out the trivial interpretation of the modal notion under consideration according to which all sentences of the form $\Box\varphi$ are true. If we seek interesting modal logics formulated in classical logic we thus need to relax the closure condition on the set IN_L. At the very least we thus need to give up the rule (Nec) or the axiom (K). Giving up (Nec) has the effect of pushing IN_L towards partial logics, since we are no longer guaranteed that all logical truth are theorems of IN_L. In contrast giving up axiom (K) opens up the possibility of the inner logic being paraconsistent, as even if $\varphi \wedge \neg\varphi \in \mathsf{IN}_L$ and $\varphi \wedge \neg\varphi \rightarrow \bot \in \mathsf{IN}_L$, we might not be able to infer $\bot \in \mathsf{IN}_L$.

However, if one pursues either strategy—(giving up (Nec) or giving up (K)—one needs to be careful not to introduce critical aspects of the strategy via the backdoor. For example, Montague (1963) shows that if one gives up (Nec) but stipulates that $\Box\varphi \rightarrow \varphi \in \mathsf{IN}_L$ paradox will persist, if IN_L is closed under the following equivalence:

$$(\star\star) \qquad (\varphi \rightarrow \neg\varphi) \equiv \neg\varphi.$$ [25]

In presenting the result we assume that a fixed-point constant δ_φ and $\varphi(\delta_\varphi)$ can be substituted in all contexts. Strictly speaking our assumptions do not license this inference since by giving up (Nec) we are no longer guaranteed that all classical tautologies are in (IN_L). However, in the first-order setting the inference is admissible, if we allow for suitable function symbols that enable us to be prove the so-called strong diagonal lemma.[26] Ultimately, in the present setting this means that we assume diagonal extensions to be closed under the following substitution schema:

$$(\mathsf{Sub}_\delta) \qquad \varphi[\delta_\psi] \leftrightarrow \varphi[\psi(\delta_\psi)].$$

Proposition 5 (Montague—in essence)**.** *Let L be a modal logic that extends classical propositional logic by* $(\star\star)$, (T) *and the axioms*

$$(\Box\star\star) \qquad \Box(\varphi \rightarrow \neg\varphi) \leftrightarrow \Box\varphi,$$
$$(\Box\mathsf{T}) \qquad \Box(\Box\varphi \rightarrow \varphi).$$ [27]

Then for all S with $L \subseteq S$, S^F is inconsistent.

Proof. Let δ be the fixed point of the formula $\Box\neg p$. Then

1. $\delta \leftrightarrow \Box\neg\delta$ — fixed-point axiom
2. $\Box\neg\delta \rightarrow \neg\delta$ — (T)
3. $\neg\delta$ — 1, 2

[25] Here, '$\equiv$' is supposed to denote that the two formulas are logically equivalent from the perspective of IN_L, that is, $\neg\varphi \in \mathsf{IN}_L$ iff $\varphi \rightarrow \neg\varphi \in \mathsf{IN}_L$. Recall that $\varphi \rightarrow \neg\varphi$ abbreviates $\neg\varphi \vee \neg\varphi$, so $(\star\star)$ ultimately amounts to conveying the idempotency of disjunction, which is assumed by basically all structural logics.

[26] For the role of function symbols and the discussion of the strong diagonal lemma see, e.g., Milne (2007); Heck (2007).

[27] $(\Box\star\star)$ forces closure of IN_L under $(\star\star)$.

4. $\Box(\Box\neg\delta \rightarrow \neg\delta)$ (□T)

5. $\Box(\Box\neg\delta \rightarrow \neg\Box\neg\delta)$ 4, (Sub$_\delta$)

6. $\Box\neg\Box\neg\delta$ 5, (□⋆⋆)

7. $\Box\neg\delta$ 6, (Sub$_\delta$)

8. δ 1, 7

□

The result shows that if modal axioms such as (T) hold for the inner logic, then even minimal closure conditions for IN_L, i.e. closure conditions that cannot be reasonably denied, will lead to paradox. Ultimately, this means that alethic modalities being factive is not itself a matter of necessity and, similarly, that an agent does not know that knowledge is factive, which seems counterintuitive to most and thus poses important limitations on formal accounts of modality.[28]

On a more positive note, one can consistently add the modal axiom D to the inner logic, that is, one can add the axiom (□D) $\Box(\Box\neg\varphi \rightarrow \neg\Box\varphi)$ to the modal logic, even if the inner logic is closed under classical consequence.

Proposition 6. *Let L be a modal logic extending classical logic by the axioms* (K), (T), (F), (4) *and* (□D) *and the rule*

(Nec$_\emptyset$) $$\frac{\varphi}{\Box\varphi}$$

whenever φ is a tautology. Then for all $S \subseteq L$, S^F is consistent.

Proof. By interpreting the modal operator logic as the truth predicate of Cantini's (1990) theory VF. □

The proposition highlights that if we let the theorems of the inner logic diverge (drastically) from those of the outer logic, we can make important steps towards more attractive formal characterizations of modal and related notions. Unfortunately, Proposition 6 is a piecemeal result and does not provide us with a general strategy for introducing formal systems for different modal notions with different modal properties. However, it is the latter that is called for and in the remainder of the chapter we try to outline such a general strategy. To this end we shall focus on the second strategy we mentioned in Section 2.1, that is, we focus on the interaction of various modal notions, provability, and truth.

[28]In some sense, Cross (2001) is thus correct in arguing that dropping K-like closure principles will not suffice for blocking paradox. We are less sure, however, whether his "Knower-plus" paradox is effective in establishing this conclusion.

Part II

The Bright Side: Steps towards a satisfactory account

So far our discussion has focused on limitative result in connection with modal, liar-like paradoxes. However, the arguably more interesting question is how these limitative results shape a formal account of modalities and propositional attitudes. We have seen that our options are limited and one might think, as many do in the case of truth, that moving to a non-classical logic is the only option. If a suitable non-classical logic is adopted, then, as in the case of truth, the naive principles of the respective notions can be non-trivially assumed. Various non-classical logics can be employed to this effect and the reader is referred to the relevant chapters of this volume or, e.g., Beall et al. (2018). In the remainder of this chapter we discuss the prospects of characterizing modality and propositional attitudes via their interaction, and, in particular, via their interaction with truth and provability. To this effect we shall conceive of modalities and propositional attitudes as predicates of sentences or structurally similar entities. There are basically two reasons for this: for one, it is common place to conceive of truth and provability as sentential predicates and by conceiving of modalities as of the same grammatical category, the interaction between these notions can be smoothly characterized. For another, in Section 2.1 we have already seen that by working in the first-order setting in which modal notions are treated as predicates, we gain further expressive strength in comparison to the operator setting, that is, unless we add further resources to the operator setting. In developing a formal account we wish to fully exploit the strength of the first-order setting rather than limit it due to certain ad hoc restrictions. Treating modalities and propositional attitudes as predicates seems appropriate for this purpose.

4 Provability-based approaches

One strategy that has been explored is to characterize various notions in relation to the notion of provability. In particular, there seems to be a close connection between belief and provability, and necessity and provability. McGee (1991) has even gone as far as arguing that logical necessity *coincides* with provability. Others have likened, at least implicitly, the notions of an agent's belief to provability in some formal system, see, e.g., Perlis (1988) below. However, if one wishes to provide a more nuanced approach regarding the interaction of provability and some other notion, a problem arises: a natural thought would be to characterize, say, logical necessity in some theory by appealing to provability in that very theory. For example, one could entertain the idea that the rule of necessitation ($\mathsf{Nec}_{\mathrm{N}}$), should be stated by means of an object-linguistic principle relating logical necessity and provability:

(Pr-Nec) $$\mathsf{Pr}(\ulcorner\varphi\urcorner) \rightarrow \mathrm{N}\ulcorner\varphi\urcorner.$$

However, if (Pr-Nec) is conceived of as an axiom of a theory Σ, then the provability predicate Pr cannot be defined by appeal to the primitive recursive relation 'is a Σ-proof of', as the latter relation will depend on the axiom (Pr-Nec). In other words, a *simultaneous axiomatization* of provability and logical necessity (or some other modality) seems impossible.

Interestingly, as developed in Niebergall (1991), one can, at least in parts, overcome this limitation by considering theory-hierarchies. In a nutshell, the idea is to start with a theory, say, PA together with (Pr_σ-Nec), i.e., (Pr-Nec) formulated using the provability predicate Pr_σ and where σ is a formula binumerating the axioms of PA. Call this theory Σ_0. Then set

$$\begin{aligned}\Sigma_{n+1} &:= \Sigma_n + (\mathsf{Pr}_{\sigma_n}\text{-Nec});\\ \Sigma_\omega &:= \bigcup_{n\in\omega}\Sigma_n.\end{aligned}$$

As Niebergall shows we can define a suitable provability predicate $\mathsf{Pr}_{\sigma_\omega}$ for Σ_ω such that the theory proves

$$(\mathsf{Pr}_{\sigma_\omega}\text{-Nec}) \qquad \mathsf{Pr}_{\sigma_\omega}(\ulcorner\varphi\urcorner) \longrightarrow \mathrm{N}\ulcorner\varphi\urcorner.$$

as well as Löb's three derivability conditions for $\mathsf{Pr}_{\sigma_\omega}$. We have found a theory in which a reasonable provability predicate of the theory and the modal predicate interact in a desirable way: a simultaneous axiomatization of provability and logical necessity has been achieved. While the theory Σ_ω may not be an attractive theory, the strategy generalizes to all theories for which the base of the progression can be proved consistent and the discussion thus serves as a prolegomenon to constructing simultaneous characterizations of provability and various modalities.[29]

We will now hint at Niebergall's construction, but readers not interested in technicalities should feel free to skip the next two paragraphs. We need to say a little bit more about hierarchies and progressions of theories. For in depth discussion of such theory-progressions we refer the reader to Feferman (1962) or Beklemishev (1999). To carry out Niebergall's construction one needs to quantify over ordinals within the arithmetic theory, that is, one has to work with an elementary ordinal notation system. Using the diagonal lemma one can then obtain an elementary two-place formula σ such that

$$\sigma(\alpha, x) \leftrightarrow \sigma_0(x) \vee \exists\beta \prec \alpha(\mathsf{Inst}(\ulcorner\mathsf{Prf}_{\sigma_0}(\ulcorner\sigma(v,x)\urcorner(\dot{\beta}/v)\urcorner, x).$$

Inst is an elementary formula saying that *x is an instance of* (Pr-Nec) *relative to the provability predicate specified by the first argument place*; $\mathsf{Prf}_\sigma(v, x)$ is a formula saying that '*x is provable in* Σ *from the set of sentences* v and α and β are variables ranging over ordinal codes. Hence, $\sigma(\alpha, x)$ numerates the axioms of Σ_α for all α.

Now, conflating the distinction between ordinals and their notation, we may define a provability predicate for Σ_ω as follows

$$\mathsf{Pr}_{\sigma_\omega}(x) :\leftrightarrow \exists y \prec \omega(\mathsf{Prf}_{\sigma_0}(\ulcorner\sigma(\dot{y}, x)\urcorner),$$

for which the desirable properties can be proved: for one, the construction guarantees that ($\mathsf{Pr}_{\sigma_\omega}$-Nec) is provable in Σ_ω and, for another, by carefully analysing the definition of $\mathsf{Pr}_{\sigma_\omega}$ one can also show it satisfies Löb's three derivability conditions.[30]

[29] We refer the reader to Niebergall (1991) for some case studies.
[30] See Niebergall (1991) for details.

Admittedly, the strategy does not provide us with a general outline of how to avoid the modal paradoxes: satisfactory assumptions on behalf of N and its interaction with provability frequently lead to inconsistent theories. For example, if one assumes that the inner logic of N is consistent, then (Pr-Nec) would lead to a contradiction due to Gödel's second incompleteness theorem. In practice, most theorists have appealed to provability predicates of subtheories in characterizing modal or propositional attitude predicates, that is, modal attitudinal and modal predicates are characterized with respect to provability in some base theory. We discuss one such example due to Perlis (1988).

Definition 7 (Perlis' theory)**.** *Let Σ be a syntax theory and $\mathcal{L}^+$ a language extending the language of Σ by the predicates* B, K, T. KFK *be the extension of* KF *in $\mathcal{L}^+$ by Löb's derivability conditions holding for* B *and the axiom*

(Know) $$\forall x(\mathsf{Sent}_{\mathcal{L}^+}(x) \longrightarrow (\mathrm{K}x \leftrightarrow \mathrm{B}x \wedge \mathrm{T}x)).$$ [31]

The consistency of KFK can be readily established by interpreting the theory in a definitorial extension of KF in $\mathcal{L}_{\mathrm{T}}$. To see this let σ be a formula representing the axioms of KF and set $\mathrm{B}(x) := \mathsf{Pr}_\sigma(x)$ and $\mathrm{K}(x) := \mathsf{Pr}_\sigma(x) \wedge \mathrm{T}x$ and, using the recursion theorem, define a regular translation function $\tau : \mathsf{Sent}_{\mathcal{L}^+} \longrightarrow \mathcal{L}_{\mathrm{T}}$ such that

$$\begin{aligned}\tau(\mathrm{T}t) &:= \mathrm{T}\tau^{\bullet}(t)\\ \tau(\mathrm{B}t) &:= \mathsf{Pr}_\sigma(\tau^{\bullet}(t))\\ \tau(\mathrm{K}t) &:= \mathsf{Pr}_\sigma(\tau^{\bullet}(t)) \wedge \mathrm{T}\tau^{\bullet}(t).\end{aligned}$$ [32]

It is then straightforward to check that for all $\varphi \in \mathsf{Sent}_{\mathcal{L}^+}$

$$\mathsf{KFK} \vdash \varphi \Longrightarrow \mathsf{KF} \vdash \tau(\varphi),$$

that is, KFK is consistent.

One may (and should) of course question plausibility of the (Know)-axiom, as it is somewhat of the mark if contemporary epistemology is to be believed. However, the approach raises a number of interesting aspects. Most importantly, in reviewing the various paradoxes we noticed that factivity, i.e., axiom schema (T) played an important role in the derivation of most paradoxes. However, against the backdrop of a theory of truth factivity (for knowledge) seems to be appropriately spelled out by

$$\forall x(\mathsf{Sent}(x) \longrightarrow (\mathrm{K}x \longrightarrow \mathrm{T}x)),$$

which, depending on the chosen theory of truth, may be consistent with the rule of necessitation for K. In other words spelling out factivity by appeal to the truth predicate blocks the derivation of Montague's theorem.

[31] To be precise Perlis does not work with the underlying truth theory KF, but the theory axiomatized by the schema

$$\mathrm{T}\ulcorner\varphi\urcorner \leftrightarrow \varphi^*$$

where the $*$-function makes sure that every negative occurrence of T in φ is turned into a positive one by pushing the negation into the scope of the truth predicate. This trick is due to Gilmore and further discussed in Feferman (1984).

[32] $\tau^{\bullet}$ represents the translation function τ.

5 Truth-based approaches

Montague's theorem and its variants appealed to the factivity of the modal notion under consideration. In modal operator logic factivity is typically spelled out via the schema (T), i.e.,

$$\Box\varphi \longrightarrow \varphi.$$

Applied to, e.g., logical necessity or knowledge factivity is arguably best spelled out by the thesis that whatever is necessary (known) is true. As mentioned, in a language with sufficient expressive resources this would seem best formalized by

$$(\mathsf{T_N}) \qquad \forall x(\mathsf{Sent}(x) \longrightarrow (\mathrm{N}x \longrightarrow \mathrm{T}x))$$

On this formulation of factivity whether paradox arises depends on the properties one ascribes to the truth predicate, that is, on the theory of truth one assumes. If the truth predicate is assumed to be naive, then paradox arises. However, in this case it is not necessary to invoke modal notions: we know that the naive truth predicate falls prey to the Liar paradox. Indeed, this might tempt one towards the view that the various paradoxes are in fact just different manifestations of the Liar paradox. The view is nicely summed up by Horsten (2002):

> "There is a view which holds that the relation between these paradoxes [the modal paradoxes, J.S.] goes deeper. The idea is that there is an underlying *conceptual connection* between the notions [the modal notions; J.S.] involved which explains *why* they are all paradoxical. This view is usually combined with the belief that at bottom there is only *one* paradoxical concept: truth. All the Liar-like paradoxes are just manifestations of the paradoxicality of the concept truth." (Horsten, 2002, p. 215; the italics are due to Horsten)

The fact that there is a prima facie route to "reducing" the paradoxes whose derivation depends on the factivity of the respective modal notion to the Liar paradox is not sufficient to argue for the above view. For example, we have seen that diagonal extensions of KD4 are inconsistent and since factivity is not assumed for these systems, it is unclear how the paradoxicality of truth could be viewed to be at the source of this inconsistency. Yet, it is not only factivity that is arguably best expressed by appealing to the notion of truth, but also the rule of necessitation. In reaction it does not seem too far fetched to understand the rule of necessitation as saying that we may infer that a sentence is necessary, once we established its truth. This would lead to the following rule of proof:

$$(\mathsf{T\text{-}Nec}) \qquad \frac{\mathrm{T}t}{\mathrm{N}t}$$

If the principles of KD4 are assumed but the rule of necessitation is replaced by (T-Nec), then inconsistency is avoided further bolstering the claim that the various paradox may reduce to the paradoxes of truth.

One caveat remains which is that the "reduction" does not seem to be neutral with respect to the particular solution to the paradoxes of truth assumed, that is, to the specific theory of truth assumed. To see this, consider the modal principles of KD4 together with the rule (T-Nec) in lieu of the standard rule of necessitation and the truth predicate characterized by the theory of truth FS or McGee's (1991) theory of non-determinate truth. Then paradox will reappear:

1.	$\gamma \leftrightarrow \neg \mathrm{N}\ulcorner \gamma \urcorner$	diagonal lemma
2.	$\mathrm{T}\ulcorner \gamma \leftrightarrow \neg \mathrm{N}\ulcorner \gamma \urcorner\urcorner$	1, FS
3.	$\mathrm{N}\ulcorner \gamma \leftrightarrow \neg \mathrm{N}\ulcorner \gamma \urcorner\urcorner$	2, (T-Nec)
4.	$\mathrm{N}\ulcorner \gamma \urcorner \rightarrow \mathrm{N}\ulcorner \neg \mathrm{N}\ulcorner \gamma \urcorner\urcorner$	3, K
5.	$\mathrm{N}\ulcorner \gamma \urcorner \rightarrow \neg \mathrm{N}\ulcorner \mathrm{N}\ulcorner \gamma \urcorner\urcorner$	4, D
6.	$\mathrm{N}\ulcorner \gamma \urcorner \rightarrow \neg \mathrm{N}\ulcorner \mathrm{N}\ulcorner \gamma \urcorner\urcorner \wedge \mathrm{N}\ulcorner \mathrm{N}\ulcorner \gamma \urcorner\urcorner$	5, 4-Ax.
7.	$\neg \mathrm{N}\ulcorner \gamma \urcorner$	6
8.	γ	1, 7
9.	$\mathrm{T}\ulcorner \gamma \urcorner$	8, FS
10.	$\mathrm{N}\ulcorner \gamma \urcorner$	9, (T-Nec)

Upon reflection paradox resurfaces because the 4-axiom leads to a semantic ascent which is not licensed by the truth theory FS. In FS semantic ascent and descent are only permissible via rules of proof. There are no axioms that enable us to introduce or eliminate the truth predicate and the addition of any principle to this effect will lead to inconsistency. The moral of this observation is that if one tries to reduce the various modal paradoxes to the paradoxes of truth and wishes to remain neutral as to the correct "solution" to the paradoxes of truth, then the axioms and rules characterizing the modality at stake should not lead to any semantic ascent or descent independent of the truth predicate. In some way, this idea squares nicely with disquotational views of truth which take disquotational function to be the principal raison d'être of the truth predicate. Taking this view a little further one could argue that modal predicates do not and should not have a disquotational function, rather that function should be reserved to the truth predicate. If such a (largely) disquotational view of truth is accepted, there is also a nice way of how axioms like the 4-axiom can be formulated in a way that conforms to the general policy:

$$(4_{\mathrm{T}}) \qquad \forall x(\mathsf{Sent}(x) \rightarrow (\mathrm{T}\ulcorner \mathrm{N}\dot{x} \urcorner \rightarrow \mathrm{N}\ulcorner \mathrm{N}\dot{x}\urcorner)).$$

In (4_{T}) the truth predicate has been introduced to the antecedent to avoid a disquotational ascent from antecedent to consequence. In more general terms the strategy is thus to avoid any form of semantic ascent or descent by introducing, if required, the disquotational truth predicate at appropriate points to the modal principles. It turns out that basically all axioms known from modal logic can be formulated in a way that avoids any form of semantic ascent and descent along the lines of the above example. While it is of course difficult to quantify over all possible theories of truth, it seems that the resulting modal principles can then be combined with most extant theories of truth: given a solution to the paradoxes in form of a truth theory, there is a straightforward strategy for extending the theory to a theory of truth and the respective modal notion. For example, Stern (2016, 2014a,b) shows how this strategy can be used to build *modal extensions* of the truth theories FS and KF and Nicolai (2017) uses Stern's strategy to introduce a modal extension of VF.[33]

Not all modal axioms incorporate a form of semantic ascent. As we have seen, some impose general conditions on the internal logic of the modality, i.e., the logic that holds in the scope of the modality under consideration. For example, the modal axiom (D) imposes the internal logic to be consistent and the axiom

$$(\mathsf{N}\text{-}\mathsf{T}_{\mathrm{T}}) \qquad \mathrm{N}\ulcorner \mathrm{N}\ulcorner \varphi \urcorner \rightarrow \mathrm{T}\ulcorner \varphi \urcorner\urcorner$$

forces the internal logic to be factive (on the classical reading).[34] In constructing modal extensions of axiomatic theories of truth one needs to be careful that the internal logic of the modal predicate matches that of the truth predicate for otherwise inconsistency may reappear.[35] However, if the inner logic of the necessity predicate and the truth predicate are compatible, then consistent modal extensions of the theory of truth are available along the strategy outlined above.[36] And this seems to amount to a substantial step in direction of satisfactory accounts of modality in expressively rich frameworks.

5.1 Semantics

The idea that the modal paradoxes can, in some sense, be reduced to the paradoxes of truth can also be nicely spelled out from a semantic perspective: suppose a given way of defining truth, say, Kripke's theory of truth. Then one can relativize the truth definition to a modal frame and use it to define an interpretation of the modal predicate, namely, as the intersection of the interpretation of the truth predicate at all accessible worlds.[37] Moreover, it turns out that if an axiomatic theory of truth "matches" a given semantic theory, then the modal extension of the axiomatic truth theory will match the modalized version of the semantic theory.

The strategy of constructing such a modal semantics is fairly simple and consists of combining the truth theory with standard possible world semantics. We now explain the construction of such a semantics in abstract terms. For applications of this strategy to particular truth theories such as Kripke's theory or the Revision theory we refer the reader to (Stern, 2016, Ch. 4). A semantic theory of truth consists of a definition of the extension of the truth predicate relative

[33]The strategy has also been used and modified by Campbell-Moore (2015a, 2016) to construct a theory of truth and self-referential probability. An approach very similar to the one discussed here was independently developed by Koellner (2016), who proposes an extension of Feferman's theory of truth DT by axioms for an absolute provability (or knowledge) predicate K.

[34]$(\mathsf{N}\text{-}\mathsf{T}_{\mathrm{T}})$ is an adaptation of ($\Box$T) that assumes our revised formulation of factivity.

[35]For example, if one assumes the axiom (D) together with the rule (T-Nec), but works with a truth predicate of KF whose internal logic is assumed to be the paraconsistent logic LP, then inconsistency will arise. Similarly, if the inner logic of the truth predicate is assumed to be strong Kleene logic one can prove that the Liar sentence is neither true nor false, then $(\mathsf{N}\text{-}\mathsf{T}_{\mathrm{T}})$ together with factivity will lead to inconsistency.

[36]Interestingly, if the negation of the inner logic of the respective truth predicate is not classical (exclusion) negation, then the notion of possibility cannot be defined on the basis of the necessity predicate. Rather one needs to introduce it as a primitive, i.e., one needs to introduce a primitive possibility predicate to the language and extend the modal theory by suitable axioms. See Stern (2014b, 2016) for details.

[37]In essence, constructions of this kind have been carried out by, e.g., Asher and Kamp (1989); Gupta and Belnap (1993); Halbach et al. (2003); Stern (2014a, 2016) for the revision theory and Halbach and Welch (2009); Caie (2012); Stern (2014b, 2016, 2015); Jerzak (2019); Field (2021) for Kripke's theory of truth.

to some ground model of the language without the truth predicate. Of course, due to Tarski we know that arithmetical truth cannot be defined in first-order arithmetics, that is, arithmetical truth is, in this sense, not first-order definable.[38] As a consequence, we will need to make use of set quantifiers in defining truth and our truth definitions will (typically) be carried out in fragments of second-order arithmetics. In the following we ignore these complications and simply identify a theory of truth with some definiens (formula) Φ such that, relative to a ground model M, for all sentences φ:

$$(\mathsf{Def}_{\mathrm{T})}) \qquad \varphi \in \|\mathrm{T}\|^{M} \ :\leftrightarrow \Phi(\varphi, M)$$

Let a frame $F = (W, R)$ consist of a nonempty set of worlds W and an accessibility relation $R \subseteq W \times W$. It is convenient to think of the worlds in W as ground models that agree on the interpretation of the syntax theory, but that may vary with respect to the interpretation of the contingent vocabulary of the language. We may then define the interpretation of the truth predicate at a world (ground model) relative to a modal frame, i.e., truth will be defined in the parameter F. This means that the definiens Φ will have an additional argument place for the frame F:

$$\varphi \in \|\mathrm{T}\|^{w} \ :\leftrightarrow \Phi(\varphi, w, F)$$

The definiens Φ yields the interpretation of the truth predicate at a world w relative to a frame F. Indeed, later it will be helpful to view Φ as defining a function f_{T} that, if applied to a world, yields the interpretation of the truth predicate at that world, e.g., $f_{\mathrm{T}}(w) = \|\mathrm{T}\|^{w}$. On the basis of the definiens Φ it is also straightforward to define the interpretation of a (universal) modal predicate N matching the accessibility relation of the frame:

$$(\mathsf{Def}_{\mathrm{N})}) \qquad \varphi \in \|\mathrm{N}\|^{w} \ :\leftrightarrow \forall v \in W(Rwv \longrightarrow \Phi(\varphi, v, F)).$$

According to this definition φ is in the interpretation of the modal predicate if Φ is true of φ at every accessible world w. As it stands the definition is independent of whether the language contains a truth predicate, but if it does, we can simplify the definition as follows:

$$\varphi \in \|\mathrm{N}\|^{w} \ :\leftrightarrow \bigcap_{v \in \{u \mid Rwu\}} \|\mathrm{T}\|^{v}.$$ [39]

On both definitions the properties of the modal predicate will depend, as in possible world semantics for modal operator logic, on the properties of the accessibility relation of the frame. For example, if R is reflexive, then for all $w \in W$

$$\varphi \in \|\mathrm{N}\|^{w} \Longrightarrow \varphi \in \|\mathrm{T}\|^{w};$$

if R is transitive, then

$$\mathrm{N}\ulcorner \varphi \urcorner \in \|\mathrm{T}\|^{w} \Longrightarrow \mathrm{N}\ulcorner \varphi \urcorner \in \|\mathrm{N}\|^{w} \ (= \bigcap_{v \in \{u \mid Rwu\}} \|\mathrm{T}\|^{v}).$$

[38] We can of course define arithmetical truth in some stronger first-order set theory.

[39] As we discuss later the two ways of defining the extension of N drive the reduction of the necessity predicate to the complex predicate ‘necessarily true’ made precise by Halbach and Welch (2009) and further discussed, e.g., in Stern (2014a,b, 2016, 2015).

In sum, the modal semantics seems to generalize the truth semantics in very much the same way standard possible world semantics generalizes model-theoretic semantics for classical logic.

Some classical axiomatic truth theories nicely characterize specific semantic theories of truth. For example, the axiomatic theory of truth Kripke-Feferman (KF) can be said to "axiomatize" Kripke's semantic theory in so far as the models of KF are precisely the expansions of the ground models to the language with the truth predicate that are obtained via Kripke's theory of truth. However, in contrast to our presentation above we assume a slightly more liberal understanding of Kripke's theory of truth. In the above presentation we have implicitly assumed that a semantic truth theory defines *the* interpretation of the truth predicate relative to a ground model, but in the following we understand Kripke's theory as providing us with a set of acceptable interpretations of the truth predicate. In $\mathsf{Def}_{\mathrm{T}}$, Φ is assumed to yield the minimal set meeting certain closure conditions, i.e., the so-called minimal Kripkean fixed point. As discussed in Fischer et al. (2015), first-order axiomatic truth theories cannot force the interpretation of the truth predicate to be minimal in that sense, but relative to a ground model they can force interpretation of the truth predicate to be satisfy the aforementioned closure conditions, i.e., to be an arbitrary fixed point. More precisely, the definiens $\Phi_{\mathsf{sk}}(\varphi, M)$ of the minimal Kripke fixed point is of the form:[40]

$$\forall X(\forall\psi(\Psi(\psi, X, M) \leftrightarrow \psi \in X) \rightarrow \varphi \in X)$$

The set of Kripke fixed points, that is, the set of interpretations relative to a ground model M that are recommended by Kripke's theory, call them Fix_M is then simply the set

$$\{X \mid \forall\psi(\Psi(\psi, X, M) \leftrightarrow \psi \in X)\},$$

that is, the set of those sets of sentences that are closed under the conditions specified by the formula Ψ. We can now state the "match" between Kripke's theory of truth, called $\mathbb{N}$-categoricity in Fischer et al. (2015). Let M be ground model and D_M its domain. Then for all $S \subseteq D_M$

$$(M, S) \Vdash \mathsf{KF} \text{ iff } S \in \mathsf{Fix}_M.$$

Let us return to the modal setting. Earlier we remarked that in the modal setting the formula Φ can be seen as defining a function, which, applied to a given world, yields the interpretation of the truth predicate at that world. Viewed in this way $\Phi_{\mathsf{sk}}(\varphi, w, F)$ amounts to the formula

$$\forall f(\forall v \in W \forall\psi(\Psi(\psi, v, f, F) \leftrightarrow \psi \in f(v)) \rightarrow \varphi \in f(w))$$

As before, modal extensions of KF will not allow us to single out the smallest or minimal such function relative to a frame. Rather any fixed point of Ψ will be an acceptable interpretation of the modal predicate on the given frame F. We set

$$\mathsf{Fix}_F := \{f \mid \forall v \in W \forall\psi(\Psi(\psi, v, f, F) \leftrightarrow \psi \in f(v))\}$$

[40]We remark that the quantifier ranging over (first-order) sentences should be conceived of as a first-order quantifier ranging over (particular) objects of the domain, and that the formula Ψ does not contain any second-order quantifier.

As mentioned earlier the set Fix_F will depend on the properties of the frame and in particular the properties of the accessibility relation. Depending on these properties different modal axioms will be true at worlds of the frame and, as a consequence, depending on these properties we obtain $\mathbb{N}$-categoricity results for different modal extension of KF. For an overview we refer the reader to Stern (2014b, 2016). As an example we discuss the theory MKF which extends the basic modal theory by the axioms T_{N}, 4_{N} and

$$(\mathrm{E}_{\mathrm{N}}) \qquad \forall x(\mathsf{Sent}(x) \rightarrow (\mathrm{T}\ulcorner\neg\mathrm{N}\dot{x}\urcorner \rightarrow \mathrm{N}\ulcorner\neg\mathrm{N}\dot{x}\urcorner)).$$

Over the truth theory KF, MKF can be viewed as the counterpart of the modal logic $\mathsf{S5}$ and one can show that MKF is true exactly in the fixed point models relative to universal frames. Let F be a Universal frame. Then

$$\text{for all } w \in W(F, w \Vdash^f \mathsf{MKF}) \text{ iff } f \in \mathsf{Fix}_F.\text{[41]}$$

Since we have fixed the intended interpretation of the syntax theory the foregoing is not quite a completeness result along the lines the well-known completeness results for modal operator logic. Rather it is a result that, given an intended model of the syntax theory, the models of the modal theory are precisely those that can be obtained by relativizing the construction of the truth model to modal frames, that is, the result tells us that the intended models of the modal theory are precisely those that are obtained by generalizing the construction of truth models.

5.2 Modifying Truth

The reader may recall that the definition of the interpretation of the modal predicate was defined as the interpretation of the truth predicate at all accessible worlds. Reflecting on the truth-clauses of possible world semantics for standard modal operator logic this seems to amount to the modification of the truth predicate by the universal modality, i.e., an understanding of the predicate N as $(\lambda x.\Box \mathrm{T}x)$. Indeed this understanding was already proposed by Kripke (1975):[42]

> Now, if a necessity operator and a truth predicate are allowed we could define a necessity predicate *Nec* applied to sentence either by $\Box\mathrm{T}(x)$ or $\mathrm{Tnec}(x)$ (...) and treat it according to the possible-world scheme sketched in the preceding paragraph [the construction sketched in the preceding section, JS].(Kripke, 1975, p. 713)

[41]By $F, w \Vdash^f \mathsf{MKF}$ we denote that MKF is true at world w relative to the frame F given the function f that assigns an interpretation to the truth and an interpretation to the necessity predicate to every world. To be precise, to obtain the result for modal extensions of KF we need to introduce a primitive existential modal predicate P in addition to the universal modal predicate N: in KF-style theories negation in the scope of the truth predicate is not classical negation and behaves differently to classical negation of the "outer logic". As a consequence, we cannot define the existential modality as $\neg\mathrm{N}\dot{\neg}$ as in standard modal operator logic. If we were to define possibility in such a way its extension would contain sentences that are neither true nor false in any world, e.g., the liar or the modal liar sentence. It would not be the case that the truth of the sentence '$\ulcorner\varphi\urcorner$ is possible' at a world w implied the truth of '$\mathrm{T}\ulcorner\varphi\urcorner$' at some world accessible from world w. See again Stern (2014b, 2016) for discussion.

[42]nec represents the primitive recursive functions, which applied to (a code of) a closed term t yields (the code of) the formula Nt.

Halbach and Welch (2009) took Kripke's idea seriously and showed how modal predicates could be "reduced" to suitable modal operators and the truth predicate. In a nutshell, their contribution consists in making rigorous our observation in the previous section that in the presence of the truth predicate in the language the definiens of ($\mathsf{Def}_{\mathrm{N}}$) can be simply recast in terms of the intersection of the interpretation of the truth predicate in all accessible worlds. Halbach and Welch (2009) provide a translation function τ from a language with a modal predicate N to a language with a modal operator $\Box$ and the truth predicate T such that $\tau(\mathrm{N}t) := \Box\mathrm{T}\tau^{\bullet}(t)$, and show that, given a frame F, truth at a world is preserved under translation if the interpretation of the modal predicate is defined as in ($\mathsf{Def}_{\mathrm{N}}$).[43] If Kripke's suggestion and Halbach and Welch's reduction of modal predicates to modified truth predicates is taken seriously, we should conceive of modal predicates in (natural) language as complex predicates that arise due to the modification of the truth predicate by a modal operator, that is, the predicate 'is necessary' should be understood as 'is necessarily true'. Understanding modal predicates in this way seems to make a compelling case for thinking that "all Liar-like paradoxes are just manifestation of the concept of truth"—at least from a semantic perspective.

Let us return to axiomatic modal theories. On the face of it, since Kripke's proposal and Halbach and Welch's reduction work merely on the semantic level it remains open to which extent the proposal applies to the axiomatic setting. Of course, given the semantics for modal extensions of KF understanding the modal predicates along the lines of Kripke's proposal is permissible, but can the close tie between modal predicates and modified truth predicates also be cashed out from a more proof-theoretic perspective? It turns out that this is possible: using Halbach and Welch's translation one can reduce the modal extensions of KF to KF formulated in a modal language assuming a suitable system of modal operator logic. For example, in the previous section we argued that the theory MKF is a modal theory that characterizes universal frames as does the modal operator logic S5 and one can show that for every sentence φ of $\mathcal{L}_{\mathsf{MKF}}$:

$$\mathsf{MKF} \vdash \varphi \Longrightarrow \mathsf{KF} \vdash_{\mathsf{S5}} \tau(\varphi).^{44}$$

More generally, if an $\mathbb{N}$-categoricity result for some modal theory relative to a class of modal frames can be given, then the modal theory can be *Kripke-reduced* to the truth theory formulated in an modal operator language characterizing the class of frames (cf. Stern, 2014a,b, 2015, 2016). This seems to further corroborate the idea that characterizing modal predicates via their interaction with the truth predicate naturally leads one to understanding these predicates as complex predicates, namely, the truth predicate modified by some suitable modal adverb or some alternative modifier. In modal logic, modal operators typically operate on the sentential level,[45] but from a linguistic perspective it is not unreasonable to think that modality can also be expressed on the subsentential level (cf. Portner, 2009),e.g., within a verb or adjectival

[43]The translation function is defined by appeal to the recursion theorem such that $\tau^{\bullet}$ represents the translation function τ.

[44]Since the language of MKF contains a primitive possibility predicate the definition of τ has to supplemented by the clause $\tau(\mathsf{P}t) := \Diamond\mathrm{T}\tau^{\bullet}(t)$.

[45]This is not quite true and particularly so if one considers quantified modal operator logic. To some extent, the distinction between the *de re* and *de dicto* reading of a formula can be considered to depend on whether on thinks of the modal operator to modify a sentence or a predicate. To distinguish these two readings logicians have introduced the machinery of λ-abstraction (cf. Fitting and Mendelsohn, 1998).

phrase. In this case it would then be natural to think of modal predicates as peculiar cases of subsentential modality, i.e., cases in which the truth predicate is modified by the respective modal notion.

5.3 Attitudes and Overgeneration

The semantics for modal predicates we introduced in Section 5.1 implements truth definitions within the possible world framework. In such a semantics it is natural to think of the semantic content of a sentence as the set of possible worlds (situations) in which the sentence is (semantically) true. Now, according to most prominent semantic approaches to truth a sentence φ will be (semantically) true iff the sentence $T\ulcorner\varphi\urcorner$ will be (semantically) true, that is, on the proposed semantic picture φ and $T\ulcorner\varphi\urcorner$ will express the same semantic content. This, in turn, implies

$$(\mathrm{ID}_{\Box})\qquad \|\Box T\ulcorner\varphi\urcorner\| = \|\Box\varphi\| = \|T\ulcorner\Box\varphi\urcorner\|;$$

$$(\mathrm{ID}_{\mathrm{N}})\qquad \|\mathrm{N}\ulcorner T\ulcorner\varphi\urcorner\urcorner\| = \|\mathrm{N}\ulcorner\varphi\urcorner\| = \|T\ulcorner \mathrm{N}\ulcorner\varphi\urcorner\urcorner\|.$$

If we take $\Box$ or N to express some alethic modality, these identities highlight the need to conceive of names of expressions (constituents of propositions) as rigid designators: if $\ulcorner\varphi\urcorner$ were to denote the ψ or, say, the parenthesis-symbol at some possible world, $\Box T\ulcorner\varphi\urcorner$ could be false or even meaningless, while it remains true that it is necessary that φ. At first glance, the identities ($\mathrm{ID}_{\Box}$) and (ID_{N}) may seem odd, for it appears to be a contingent matter that sentences have the meaning they do. The chapter is not the place to evaluate this problem at length, but suffices it to say that according to disquotationalist like Field (1994) the English truth predicate applies to sentences as understood in English when uttered (in a given context), and we take it that proponents of propositional truth understand the truth predicate in a similarly way. Such views have precisely the effect of treating a name $\ulcorner\varphi\urcorner$ rigidly. In other words the disquotationalist commits to a *de re*-reading of $\Box T\ulcorner\varphi\urcorner$. Importantly, in natural language our understanding of the truth predicate frequently aligns with the *de re*-reading and thus justifies the above identification.[46]

If one takes possible world semantics to be an apt semantics for propositional attitude reports roughly along the line of Hintikka (1962) and reads ($\mathrm{ID}_{\Box}$) and (ID_{N}) with a notion like belief in mind, then these identities are troublesome. Clearly, it is possible to believe that every even number is the sum of two primes without believing Goldbach's conjecture true and *vice versa*; it is possible to believe that Superman is strong without believing that 'Clark Kent is strong' is true. Arguably, this was the reason why Kripke suggested that defining attitudinal predicates such as belief using the complex predicate $T\ulcorner\Box\dot{x}\urcorner$ as opposed to $\Box Tx$: the attitudinal predicate is defined as $\ulcorner$*... believes ...*$\urcorner$ *is true* and not as *... believes-true '... '* (cf. Kripke, 1975, p. 713, Fn 33). While on this definition the mentioned linguistic counterexamples might be avoided, the definition does not solve the formal challenge arising in possible world semantics: the problem is precisely that both potential definiens are equivalent on this semantics. Stern

[46]It is another question whether our understanding of truth *always* aligns with the *de re*-reading. However, if syntax is not understood rigidly, then a formal account of the interaction of truth and modality (operator or predicate) seems very much out of reach. We refer to Wallace (1970); Thomason (1976); Peacocke (1978); Gupta (1978); Davies (1978) for some discussion. See also Heck (2021) for a discussion of related themes.

(2021) suggests that these counterexamples are a variation on Soames's (1987) arguments of why semantic content cannot be sets of truth supporting circumstances. Soames concludes that semantic content should be analysed in terms of structured proposition. On Soames's view propositional attitude verbs express relations between an agent and a structured proposition. The analysis suggests conceiving of attitudinal predicates as primitive predicates rather than defined (explicitly or implicitly) using a sentential operator and the truth predicate. The view would still square well with the strategy of building axiomatic theories of the respective notions by exploiting their interaction with the notion of truth.[47] But the semantics should at best be understood instrumentally and the existence of *Kripke-reductions* that conceive of the modal predicates as modified truth predicates should be understood as an artefact of the semantics.

Typically proponents of truth-conditional (possible world) semantics resist Soames's conclusion by taking the attitude contexts to be representation sensitive.[48] There are (at least) two options how this can be implemented (cf. Stern, 2021). The first option takes the semantics of attitude reports to be non-compositional: the semantic value (content) of the attitude report does not solely depend on the value of its components, but also on its representation. Even though $\|\varphi\| = \|\mathrm{T}\ulcorner\varphi\urcorner\|$ we cannot conclude $\|\Box\varphi\| = \|\Box\mathrm{T}\ulcorner\varphi\urcorner\|$ because φ and $\mathrm{T}\ulcorner\varphi\urcorner$ represent the same content in different ways. Accounts of this type can be traced back to Carnap (1947) and his notion intensional isomorphism, and have been embraced by, e.g., Lewis (1970); Kratzer (2022). On the second option the relations in which agents stand to some particular content will depend on the specific context of the attitude report.[49] On such a semantics one may believe $\|\varphi\|$, but not believe $\|\mathrm{T}\ulcorner\varphi\urcorner\|$ as a change in context may lead to a change of the belief-relation, that is in the present setting, a change of which worlds are doxastically accessible. Compositionality is retained by introducing a further semantic parameter into the logical form of attitude reports. If possible world semantics is modified along the lines of either option the strategy for giving a semantics for attitudinal predicates can be employed without making $(\mathrm{ID}_\Box)$ and (ID_N) true. As suggested by Kripke an attitudinal predicate $\mathrm{N}x$ should then be defined (explicitly or implicitly) by the formula $\mathrm{T}\ulcorner\Box\dot{x}\urcorner$ (and not $\Box\mathrm{T}x$) to avoid the equivalence between believing and believing-true.

We argued the structured propositions theorists should best conceive of attitudinal (and arguably) modal predicates as primitive predicates, that is, they should not think of these predicates as defined via the truth predicate and a sentential operator. In contrast, it seems advantageous for proponents of (intensional) truth-conditional semantics to conceive of such

[47]Unfortunately things are slightly more subtle than that and the view will fit better with some theories than with other. For example, in constructing a modal extension of FS Stern (2014a; 2016) introduces the axiom

$$\forall x(\mathsf{Sent}(x) \rightarrow (\mathrm{T}\ulcorner\mathrm{N}\dot{x}\urcorner \leftrightarrow \mathrm{N}\ulcorner\mathrm{T}\dot{x}\urcorner)),$$

which is not acceptable for obvious reasons. In contrast, modal extensions of KF are based on the axiom

$$\forall x(\mathsf{Sent}(x) \rightarrow (\mathrm{T}\ulcorner\mathrm{N}\dot{x}\urcorner \leftrightarrow \mathrm{N}x),$$

which does not seem to lead to counterintuitive consequences.

[48]Alternatively, one can also follow, e.g., Moltmann (2020) and resist the relational account of propositional attitudes.

[49]The idea is to implement contextualist ideas à la Crimmins and Perry (1989) within possible world semantics for attitude reports. See Stern (2021) for more details.

predicates along the lines of the Kripke-reduction. Understanding modal and attitudinal predicates as suitable complex predicates neatly explains how to make sense of attitudinal and modal predicates applying to linguistic material such as sentences. It also provides us with a straightforward account of quantification across the argument positions of various modal and attitudinal predicates. If such predicates are understood as primitives one may think that different attitudinal predicates will apply to different attitudinal objects (Prior, 1971; Vendler, 1967) and the question then arises of how quantified statements connecting various modal and attitudinal notions ought to be understood.

Summing up, tying formal approaches of modal notions to formal theories of truth, and to the interaction of the modal predicates and the truth predicates leads to workable and, arguably, satisfactory formal characterizations of modality in expressively rich frameworks. The strategy can be seen as "reducing" the modal paradoxes to the paradoxes of truth: given a consistent theory of truth we can construct consistent modal theories. Interestingly, this "reduction" in some sense vindicates orthodox modal operator logic, as it turns out that these modal predicates can be understood as modified truth predicates defined in modalized versions of the truth theory under consideration. Whether the understanding of modal predicates as modified truth predicates is taken seriously will ultimately depend on ones semantic outlook.

6 Further Issues

In this chapter we aimed to introduce the modal paradoxes and to point to some strategies for answering the challenges they pose for satisfactory accounts of modality and propositional attitudes. One central tenet underlying our discussion was that modal paradoxes and related phenomena arise independently of whether modal notions are conceived of as predicates or sentential operators. Indeed, this idea is supported by the paradoxes of indirect discourse (Prior, 1961, 1971). In formal epistemology these paradoxes are sometimes discussed under the label of Brandenburger-Kiesler paradox (Brandenburger and Keisler, 2006).[50] In essence, these paradoxes show that if propositional modal logic is enriched by sentential quantification, then the paradoxes will resurface. In this setting sentential quantification amounts to a restricted form of second-order quantification and the paradoxes cannot be straightforwardly answered by invoking a non-naive theory of truth. Rather in this setting it seems that one has to revise the standard rules of quantification and, arguably, move to some form free logic of propositional quantification (see, e.g., Asher, 1990; Bacon et al., 2016; Bacon, 2021). Such an approach to modal paradox and the paradoxes of indirect discourse may be preferable to proponents of *Higher-Orderism* (see, e.g. Fritz and Jones, 2023). In contrast, if quantification over sentences is conceived of as first-order quantification, then paradoxes of indirect discourse ultimately amount to variant of the *Epimenides paradox* and can be addressed along the lines of the strategy discussed in Section 5. To our mind, this a more natural and intuitive approach, but of course opinions will differ.

In the spirit of full disclosure we end this chapter by outlining a problem that arises if we combine truth theories with possible world semantics, which has recently been discussed

[50]The Brandenburger-Kiesler is a theorem about possible world semantics, whereas Prior's paradoxes of indirect discourse are not directed towards a specific semantics.

by Halbach (2021) and labeled *the fourth grade of modal involvement.* In our presentation we implicitly assumed that we possess names for all object in the domain. In absence of this idealizing assumption it is natural to conceive of the modal predicate, whether primitive or complex, as a two-place modal satisfaction predicate which applies to a formula and a finite sequence of objects of the domain. Perhaps surprisingly, this creates a problem for proponents of actualist quantification, as they are forced to countenance the existence of sequents of non-existent objects, which will make everyone save the hardcore Meinongean feel uneasy. No such problem arises for proponents of possibilist quantification and perhaps the moral of the argument is that one needs to assume a possibilist or, following Williamson (2013), a necessitist position regarding the metasemantic vocabularly, but further research is required to this effect.

Acknowledgements Work on this chapter was supported by the ERC Starting Grant *Truth and Semantics* (TRUST, Grant no. 803684). I wish to thank Lorenzo Rossi for the invitation to contribute to this volume and Carlo Nicolai, Poppy Mankowitz, Simone Picenni, Tedy Nenu, Will Stafford, and Xinhe Wu for very helpful comments on an earlier draft of this chapter. I also wish to thank the audience of a talk in Warsaw on related material for helpful feedback.

References

Asher, N. (1990). Intentional paradoxes and an inductive theory of propositional quantification. In Parikh, R., editor, *Theoretical Aspects of Reasoning about Knowledge*. Morgan Kaufmann.

Asher, N. and Kamp, H. (1989). Self-reference, attitudes and paradox. In Chierchia, G., Partee, B. H., and Turner, R., editors, *Properties, Types, and Meaning. Vol. I: Foundational Issues*, pages 85–158. Kluwer.

Bacon, A. (2021). Opacity and paradox. In Nicolai, C. and Stern, J., editors, *Modes of Truth: The Unified Approach to Modality, Truth, and Paradox*, pages 151–181. Routledge.

Bacon, A., Hawthorne, J., and Uzquiano, G. (2016). Higher-order free logic and the prior-kaplan paradox. *Canadian Journal of Philosophy*, 46(4-5):493–541.

Beall, J., Glanzberg, M., and Ripley, D. (2018). *Formal theories of truth.* Oxford University Press.

Beklemishev, L. (1999). Proof-theoretic analysis by iterated reflection. *Archives of Mathematical Logic*, 42(6):515–552.

Blackburn, P., de Rijke, M., and Venema, Y. (2001). *Modal Logic.* Cambridge University Press, Cambridge.

Bonnano, G. (1996). On the Logic of Common Belief. *Mathematical Logic Quarterly*, 42:305–311.

Boolos, G. S. (1993). *The Logic of Provability.* Cambridge University Press.

Brandenburger, A. and Keisler, H. J. (2006). An impossibility theorem on beliefs in games. *Studia Logica*, 84:211–240. Special Issue Ways of Worlds II. V. F. Hendricks and S. A. Pedersen, Editors.

Caie, M. (2012). Belief and Indeterminacy. *Philosophical Review*, 121(1):1–54.

Caie, M. (2013). Rational Probabilistic Incoherence. *Philosophical Review*, 122(4):527–575.

Campbell-Moore, C. (2015a). How to express self-referential probability. a Kripkean proposal. *The Review of Symbolic Logic*, 8(4):680–704.

Campbell-Moore, C. (2015b). Rational Probabilistic Incoherence? A Reply to Michael Caie. *The Philosophical Review*, 124(3):393–406.

Campbell-Moore, C. (2016). *Self-referential probability*. PhD thesis, Ludwig-Maximilians-Universität München.

Cantini, A. (1990). A theory of formal truth arithmetically equivalent to ID_1. *The Journal of Symbolic Logic*, 55:244–259.

Carnap, R. (1947). *Meaning and Necessity*. University of Chicago Press, Chicago.

Crimmins, M. and Perry, J. (1989). The prince and the phone booth: Reporting puzzling beliefs. *The Journal of Philosophy*, 86(12):685–711.

Cross, C. B. (2001). The paradox of the knower without epistemic closure. *Mind*, 110:319–333.

Czarnecki, M. and Zdanowski, K. (2019). A modal logic of a truth definition for finite models. *Fundamenta Informaticae*, 164(4):299–325.

Davies, M. K. (1978). Weak necessity and truth theories. *The Journal of Philosophical Logic*, 7:415–439.

Egré, P. (2005). The knower paradox in the light of provability interpretations of modal logic. *Journal of Logic, Language and Information*, 14:13–48.

Fagin, R., Halpern, J. Y., Moses, Y., and Vardi, M. Y. (1995). *Reasoning about Knowledge*. MIT Press.

Feferman, S. (1962). Transfinite recursive progression of axiomatic theories. *The Journal of Symbolic Logic*, 27(3):259–316.

Feferman, S. (1984). Toward useful type-free theories.i. *The Journal of Symbolic Logic*, 49(1):75–111.

Field, H. (1994). Deflationist views of meaning and content. *Mind*, 103:249–285.

Field, H. (2021). Properties, propositions, and conditionals. *Australasian Philosophical Review*, forthcoming.

Fischer, M., Halbach, V., Kriener, J., and Stern, J. (2015). Axiomatizing semantic theories of truth? *The Review of Symbolic Logic*, 8(2):257–278.

Fitting, M. and Mendelsohn, R. L. (1998). *First-Order Modal Logic*. Kluwer Academic Publishers.

Friedman, H. and Sheard, M. (1987). An axiomatic approach to self-referential truth. *Annals of Pure and Applied Logic*, 33:1–21.

Fritz, P. and Jones, N. K., editors (2023). *Higher-Order Metaphysics*. Oxford University Press. forthcoming.

Germano, G. (1970). Metamathematische Begriffe in Standardtheorien. *Archiv für mathematische Logik und Grundlagenforschung*, 13:22–38.

Gupta, A. (1978). Modal logic and truth. *Journal of Philosophical Logic*, 7:441–472.

Gupta, A. (1982). Truth and paradox. *Journal of Philosophical Logic*, 11:1–60.

Gupta, A. and Belnap, N. (1993). *The revision theory of truth*. The MIT Press.

Halbach, V. (2006). How not to state T-sentences. *Analysis*, 66:276–280. Correction in Analysis 67: 268.

Halbach, V. (2008). On a side effect of solving Fitch's paradox by typing knowledge. *Analysis*, 68:114–120.

Halbach, V. (2014). *Axiomatic Theories of Truth, 2nd Edition*. Cambridge University Press.

Halbach, V. (2021). The fourth grade of modal involvement. In Nicolai, C. and Stern, J., editors, *Modes of Truth: The Unified Approach to Modality, Truth, and Paradox*, pages 209–230. Routledge.

Halbach, V. and Leigh, G. (2022). The road to paradox: A guide to syntax, truth, and modality. Manuscript, Oxford and Gothenburg.

Halbach, V., Leitgeb, H., and Welch, P. (2003). Possible-worlds semantics for modal notions conceived as predicates. *Journal of Philosophical Logic*, 32:179–222.

Halbach, V. and Welch, P. (2009). Necessities and necessary truths: A prolegomenon to the use of modal logic in the analysis of intensional notions. *Mind*, 118:71–100.

Heck, R. G. (2007). Self-reference and the languages of arithmetic. *Philosophia Mathematica*, 15(3):1–29.

Heck, R. K. (2021). Disquotationalism and the Compositional Principles. In Nicolai, C. and Stern, J., editors, *Modes of Truth: The Unified Approach to Modality, Truth, and Paradox*, pages 115–150. Routledge.

Hintikka, J. (1962). *Knowledge and Belief*. Cornell University Press, Ithaca and London.

Horsten, L. (2002). An axiomatic investigation of provability as a primitive predicate. In Halbach, V. and Horsten, L., editors, *Principles of Truth*, pages 203–220. Ontos Verlag.

Horsten, L. and Leitgeb, H. (2001). No future. *Journal of Philosophical Logic*, 30:259–265.

Jerzak, E. (2019). Non-classical knowledge. *Philosophy and Phenomenological Research*, 98(1):190–220.

Kaplan, D. and Montague, R. (1960). A paradox regained. *Notre Dame Journal of Formal Logic*, 1:79–90.

Koellner, P. (2016). Gödel's disjunction. In Horsten, L. and Welch, P., editors, *Gödel's Disjunction. The Scope and Limits of Arithmetical Knowledge*, pages 148–188. OUP.

Koons, R. C. (1992). *Paradoxes of Belief and Strategic Rationality*. Cambridge University Press.

Kratzer, A. (2022). Attitude ascriptions and speech reports. In Altshuler, D., editor, *Linguistics Meets Philosophy*, pages 17–50. Cambridge University Press, New York.

Kripke, S. (1975). Outline of a theory of truth. *The Journal of Philosophy*, 72:690–716.

Kurahashi, T. (2020). Rosser provability and normal logics. *Studia Logica*, 108:597–617.

Leigh, G. E. and Rathjen, M. (2012). The friedman-sheard programme in intuitionistic logic. *The Journal of Symbolic Logic*, 77(3):777–806.

Lewis, D. (1970). General semantics. *Synthese*, 22(1/2):18–67.

Makinson, D. (1971). Some embedding theorems for modal logic. *Notre Dame Journal of Formal Logic*, 12(2):252–54.

McGee, V. (1985). How truthlike can a predicate be? A negative result. *The Journal of Philosophical Logic*, 14(4):399–410.

McGee, V. (1991). *Truth, Vagueness and Paradox*. Hackett Publishing Company, Indianapolis.

Milne, P. (2007). On Gödel Sentences and What They Say. *Philosophia Mathematica*, 15(3):193–226.

Moltmann, F. (2020). Truthmaker semantics for natural language: Attitude verbs, modals, and intensional transitive verbs. *Theoretical Linguistics*, 46(3-4):159–200.

Montague, R. (1963). Syntactical treatments of modality, with corollaries on reflexion principles and finite axiomatizability. *Acta Philosophica Fennica*, 16:153–167.

Myhill, J. (1960). Some remarks on the notion of proof. *The Journal of Philosophy*, 57:461–471.

Nicolai, C. (2017). Necessary truths and supervaluations. In De Florio, C. and Giordani, A., editors, *From arithmetic to metaphysics. A path through philosophical logic.* De Gruyter. forthcoming.

Nicolai, C. and Stern, J. (2021). The Modal Logic of Kripke-Feferman Truth. *The Journal of Symbolic Logic*, 86(1):362–395.

Niebergall, K.-G. (1991). Simultane objektsprachliche Axiomatisierung von Notwendigkeits- und Beweisbarkeitsprädikaten. Masterthesis, Ludwig-Maximilians Universität München.

Niemi, G. (1972). On the existence of a modal antinomy. *Synthese*, 23:463–476.

Peacocke, C. (1978). Necessity and truth theories. *The Journal of Philosophical Logic*, 7:473–500.

Perlis, D. (1988). Languages with self-reference II: Knowledge, belief, and modality. *Artificial Intelligence*, 34:179–212.

Portner, P., editor (2009). *Modality*. Oxford University Press, New York.

Prior, A. (1961). On a family of paradoxes. *Notre Dame Journal of Formal Logic*, 2:16–32.

Prior, A. (1971). *Objects of Thought*. Clarendon Press, Oxford.

Schwarz, W. (2013). Variations on a Montagovian theme. *Synthese*, 190:3377–3395.

Schweizer, P. (1992). A syntactical approach to modality. *Journal of Philosophical Logic*, 21:1–31.

Slater, H. (1995). Paraconsistent logics? *Journal of Philosophical Logic*, 24:451–454.

Smorynski, C. (1985). *Self-Reference and Modal Logic*. Springer Verlag.

Smorynski, C. (2004). Modal logic and self-reference. In Gabbay, D. and Guethner, F., editors, *Handbook of Philosophical Logic, 2nd Edition*, volume 11, pages 1–55. Kluwer Academish Publishers.

Soames, S. (1987). Direct reference, propositional attitudes, and semantic content. *Philosophical Topics*, 15(1):47–87.

Stern, J. (2014a). Modality and Axiomatic Theories of Truth I: Friedman-Sheard. *The Review of Symbolic Logic*, 7(2):273–298.

Stern, J. (2014b). Modality and Axiomatic Theories of Truth II: Kripke-Feferman. *The Review of Symbolic Logic*, 7(2):299–318.

Stern, J. (2014c). Montague's Theorem and Modal Logic. *Erkenntnis*, 79(3):551–570.

Stern, J. (2015). Necessities and Necessary Truths. Proof-theoretically. *Ergo*, 2(10):207–237.

Stern, J. (2016). *Toward Predicate Approaches to Modality*, volume 44 of *Trends in Logic*. Springer, Switzerland.

Stern, J. (2021). Belief, truth, and ways of believing. In Nicolai, C. and Stern, J., editors, *Modes of Truth: The Unified Approach to Modality, Truth, and Paradox*, pages 151–181. Routledge.

Stern, J. and Fischer, M. (2015). Paradoxes of Interaction? *Journal of Philosophical Logic*, 44(3):287–308.

Thomason, R. (1976). Necessity, quotation, and truth: An indexical theory. In Kasher, A., editor, *Language in Focus*, pages 119–138. D. Reidel Publishing Company.

Thomason, R. (1980). A note on syntactical treatments of modality. *Synthese*, 44:391–395.

Vendler, Z. (1967). *Linguistics in Philosophy*. Cornell University Press.

Visser, A. (1989). Semantics and the Liar Paradox. In Gabbay, D., editor, *Handbook of Philosophical Logic*, pages 617–706. Dordrecht.

Wallace, J. (1970). On the frame of reference. *Synthese*, 22(1):61–94.

Williamson, T. (2013). *Modal logic as metaphysics*. Oxford University Press, Oxford.