King's Research Portal

Link to publication record in King's Research Portal

1 A Guide to the Unified Approach to Truth, Modality, and Paradox

Carlo Nicolai and Johannes Stern

The notion of truth, modal notions, and doxastic notions such as belief and knowledge play a crucial role in contemporary philosophy. From its very beginnings philosophy has sought to fully understand these notions. Following the *linguistic turn* and the rise of analytic philosophy, the study of the uses of the aforementioned notions in (natural) language has become an important part of the data against which a theory has to be tested and sharpened. Indeed research on truth within this tradition has focused on the role and uses of the *truth predicate* in language—be it in natural language or some more regimented theoretical language. Research on belief has been importantly influenced by how *belief reports* are to be best understood and similar remarks apply to modal notions and knowledge.

In natural language we say

(1) This apple is red.
(2) Goldbach's conjecture is true.
(3) Goldbach's conjecture is necessary.
(4) Mary believes Goldbach's conjecture.

In light of these examples it seems, at least prima facie, that saying of Goldbach's conjecture that it is true, possible, or believed is not fundamentally different from the claim made in (1): in both cases we seem to ascribe a particular characteristic, feature, or property to an object, that is, a specific apple in (1) but the denotatum of Goldbach's conjecture in (2), (3), and (4). In other words 'Goldbach's conjecture' occupies a subject position in (2), (3), and (4) that typically can be occupied by first-order singular terms and can be bound by first-order quantification.[1] The objects of truth are not apples, however, but commonly thought to be sentences (types), utterances, or propositions.[2] While sentences or propositions are, of course, objects of a very different kind from apples, they remain objects of the same semantic type on this view. In the context of Montague Grammar, they will all be objects of type *e*. As a consequence, the truth predicate enables generalization over sentences or

propositions by means of standard first-order quantification. Indeed this feature is widely acknowledged as one of the principal uses of the truth predicate in language (see, e.g., Quine, 1970).

However, according to an alternative view generalization over sentences or propositions does not arise via quantification into the argument position of a sentential predicate, but rather via some sort of higher-order quantification over sentence position. On this view, in sentences like (2), despite appearances to the contrary, we do not ascribe truth to some object of type e. Rather, on this alternative semantic picture the truth predicate will not occur in the logical form of the sentence. Instead the truth predicate will be eliminated and replaced by propositional quantification, which is standardly analysed as quantification over objects of type $\langle s, t \rangle$. Arguably this line of research originates with Ramsey's Redundancy Theory of Truth (Ramsey, 1927, 1929) and finds more contemporary proponents in the form of the Prosentential Theory of Truth (Grover et al., 1975; Grover, 1992) and Mulligan (2010).

In the literature on truth the idea that the quantifier in sentences such as

(5) Every axiom of ZFC is either true or false

is a propositional, i.e., a higher-order quantifier ranging over objects of type $\langle s, t \rangle$ remains a minority position and truth is usually formalized (and interpreted) by a first-order predicate constant. In contrast, and somewhat surprisingly, in philosophical logic it is customary to formalize attitude verbs such as 'believe', as well as, knowledge and various modal notions by means of sentential operators that apply to object of type $\langle s, t \rangle$. Quantification into the argument position of attitude verbs, knowledge, and the modal notions then need to take the form of propositional quantification, that is, quantification over objects of type $\langle s, t \rangle$ (Bull, 1969; Fine, 1970).[3]

The discrepancy in the formal treatment of truth, modality, and doxastic notions leads to the problem of cross-quantification: how is the quantifier in sentences like

(6) Everything Mary believes is true

best understood? Is it a first-order quantifier as required by the standard view on truth or is it a propositional quantifier that binds the argument position of the attitude verb if the latter is formalized as a doxastic operator?

Unfortunately, sentences like (6) are abundant in natural language and, specifically, philosophical discourse, so the problem of quantification cannot be put aside. To resolve the problem, some level of uniformization between the formal treatment of truth, doxastic notions, and

modal notions seems required. In other words, the notion of truth, doxastic notions, and modal notions as they appear in sentences like (6) need to be treated as expressions of the same logical category to allow for the type of cross-quantification displayed in (6).[4]

If the grammatical form of sentences (2)–(6) is taken as a guide, attitude verbs, knowledge, and modal notions may be best conceived of as predicates. This would facilitate interpreting quantification in these sentences along the lines of first-order quantification. Alternatively, one could take a revisionist stance towards the surface-level grammatical form and conceive of quantification in these sentences as propositional quantification: there is no singular term position that is bound by the quantifier expression in (5) and (6). The quantifier ranges over sentence position rather than their nominalizations. All else equal, taking the first-order route may seem advantageous. First-order quantification is well understood and it comes with a fully developed semantics and proof theory. In contrast, while there exist attractive semantics for higher-order quantification and more specifically propositional quantification, these semantics typically cripple the language's expressive resources: self-reference is usually eliminated in favour of an open-ended hierarchy of types. As will become clear later, Bacon's *Opacity and Paradox* highlights that this need not be the case, i.e., propositional quantification can accommodate self-reference without appealing to an open-ended hierarchy of types. However, the semantics and proof theory of non-hierarchical theories of propositional quantification is underdeveloped in comparison to the first-order approach.

Putting all of this together, there is a strong case to uniformly conceive of the objects of truth, modality, knowledge, and attitudinal relations as first-order objects that can be quantified over via first-order quantification—call this the Unified Approach. However, grammatical form and a well developed proof theory and semantics do not amount to conclusive arguments for adopting the first-order approach. If convincing arguments in favour of a higher-order approach are available, the latter may well be preferable. Indeed, several philosophers and logicians pursue the higher-order approach. The chapters by Bacon and Studd in this volume contribute to the development of the higher-order approach.

Independently of whether one prefers a first-order or higher-order approach, a uniform treatment of truth, modality, and attitudinal relation seems a prerequisite for an adequate semantics and proof theory of sentences which involve cross-quantification in the sense discussed above. However, unless special precautions are taken, such a uniform treatment of the notion of truth, modal notions, and propositional attitudes will encounter paradoxes such as the liar paradox, Montague's paradox (Myhill, 1960; Montague, 1963), and paradoxes of belief (Thomason, 1980; Cross, 2001).[5] A popular reaction to the paradox is to banish self-reference from the framework by introducing either syntactic

or semantic restriction (e.g., typing restrictions) that prevent expressing self-reference of any sort. Such approaches are hardly satisfactory, as has been convincingly argued by Kripke (1975) and others: they don't provide an interpretation of fragments of natural language that fit the data. Rather than banishing self-reference from the framework, strategies have to be put in place for handling vicious or infelicitous forms of self-reference.

The paradoxes not only pose a challenge for any adequate formal account of truth and related notions, they also highlight the need of developing such an account in an unified way, that is, in a framework in which the various notions are allowed to freely interact. As has been stressed by Horsten and Leitgeb (2001), Halbach (2006, 2008), and Stern and Fischer (2015), the interaction of the various notions may trigger new and unexpected pathologies. This indicates that proceeding in a piecemeal fashion and tackling truth, modality, or the propositional attitudes individually rather than simultaneously is bound to lead to problems.[6]

In the first-order framework, self-reference is usually achieved via the various forms of nominalizations available, e.g., via Gödel numbering or other means that enable one to talk about expressions of the language (alternatively, component parts of structured propositions). However, in the higher-order framework, that is, a framework that allows for quantification into sentence-position, the liar-like paradoxes arise directly via quantification into the argument position of the truth-like notions (or related means). In this form the paradoxes are known as paradoxes of indirect discourse, and have been discussed, e.g., by Prior (1961, 1971).[7] The paradoxes of indirect discourse are basically versions of the Epimenides paradox, but in contrast to the more standard liar-like paradoxes, require rethinking the logic and semantics of propositional quantification rather than the logic and semantics of the truth-like notions. An example to this effect is given by Bacon's contribution but also Asher (1990), yet research on the semantics of propositional quantification in light of the paradoxes of indirect discourse is somewhat more scarce than the research on adequate semantics for the truth predicate and related notions. This volume aims to continue research on the advantages and limitations of the first-order framework. Addressing such challenges necessitates a closer look at the truth predicate, its role and function in language, and its semantics. And this is where the journey begins.

1.1 Truth: Semantics and Disquotation

Even if truth is best conceived as a predicate applying to first-order entities, this does not settle many other issues concerning the role of truth in language and reasoning. Theorizing about truth comes in different forms.

1.1.1 Truth and Semantics

The standard *semantic* notion of truth lives in a metalanguage that is distinct from the fragment of natural language one wants to analyse, the object-language. This metalanguage comprises a rich ontology of mathematical objects and resources to talk about the semantic values of the linguistic components of the object-language. Semantic truth is typically defined in such a metalanguage, although it is possible to employ a direct axiomatization of one's metatheoretic truth predicate. Semantic truth can be employed to study the object-linguistic truth predicate, whose properties may not coincide with the semantic one.

Several theorists interpret natural language data as supporting the desideratum that the semantic value of a sentence A ought to be the same as the semantic value of '"A" is true', where A is an arbitrary sentence of the object-language. There are different ways to construct a formal semantics with this property. The liar paradox tells us that a formal semantics in which A and '"A" is true' have the same semantic value cannot validate all principles of classical logic. Kripke (1975) is arguably the starting point of modern investigations on the semantics of self-applicable truth: the formal semantics proposed by Kripke does not validate some of the classical logical principles for negation (or, equivalently, the law of excluded middle). The sentences that generate paradox are, in Kripke's semantics, truth-value gaps.

Kripke's approach is the basis for semantic frameworks that are mathematically very close to Kripke's original semantics, but that are conceptually quite apart. By conceiving of the liar paradox as a datum supporting the inconsistency of truth, paraconsistent approaches to the semantics of the truth predicate are based on the idea that there are sentences that are both true and false (Priest, 2006; Beall, 2009), that is, truth-value gluts. Such sentences crucially contain the object-linguistic truth predicate. The paraconsistent semantics developed in Priest (2006) also restrict some principles of negation (or, equivalently, the classical inference *ex-falso quodlibet*).

A paraconsistent version of Kripke's theory of truth can be constructed to yield models of the language with a self-applicable truth predicate that validate all classical inferences—including the ones involving negation—but that invalidate some classical meta-inferences (Ripley, 2012; Cobreros et al., 2013). Let's restrict our attention to models with only three values $\{1, \frac{1}{2}, 0\}$. Fixed-point models inspired by the non-transitive approach do not validate the structural rule of *cut*. This is achieved by modifying the notion of logical consequence associated with Kripke fixed-point models. Assuming a fully structural notion of logical consequence, φ follows from Γ precisely when designated truth-values are preserved. In the non-transitive approach, φ follows from Γ precisely when, if all sentences in Γ receive value 1, then φ cannot receive value 0. This is

the notion of *strict-tolerant consequence*. Since both the liar sentence λ and its negation $\neg\lambda$ do not receive value 0 in any model, they are both valid. However, the non-transitive approach to truth does not lead to triviality, as the meta-inference of cut is not validated.

Paul Égré's *Half Truths and Liars* aims to refine the analysis of the semantics of the object-linguistic truth predicate given by the non-transitive approach. According to the standard non-transitive theorist, sentences such as λ and '2 + 2 = 4' do not differ in their strict semantic status: they are both valid. However, unlike what happens with 2 + 2 = 4, the negation of λ is also valid, but there is no sense in which the non-transitivist can deem λ to be *less true* than '2 + 2 = 4'. In the standard non-transitive approach, λ and '2 + 2 = 4' can be seen to diverge at the *pragmatic level*, in particular in their assertibility conditions. '2 + 2 = 4' is strictly and tolerantly assertible, whereas λ is tolerantly but not strictly assertible.

By reflecting on natural language data, Égré proposes a genuinely *semantic* analysis of 'is true' as a vague predicate. He analyses 'is true' as as an absolute gradable adjective. 'Is true' should be compatible with uses such as 'true in some sense', or 'true in some respect', but also faithful to absolute uses of 'is true' as 'true *simpliciter*' or 'perfectly true'. Égré defines a partial and a total meaning of 'is true', roughly corresponding to the semantic values of 'true in some respects' and 'true in all respects'. The liar sentence λ cannot be true in all respects, but can only be *half true*. Crucially, Égré's analysis reconciles the non-transitive approach to truth with its original analysis of vagueness as an essentially semantic phenomenon. This shows that the pragmatic machinery of assertibility conditions applied to the analysis of truth ascriptions may not be intrinsic to the non-transitive approach.

1.1.2 Truth and Logic

Semantics is one theoretical context in which the notion of truth has been employed. Traditional truth-theoretic deflationism denies that it is an important one. Truth is best conceived of as a quasi-logical device that supports correct reasoning by enhancing the expressive capabilities of our language. This *logical* notion of truth holds firm some suitable formulation of the T-schema: '"A" is true' is equivalent to A.[8] It investigates consistent (non-trivial) ways of characterizing truth by means of suitable quasi-logical principles. Such principles typically need to (i) support the expressive power afforded by the T-schema, and (ii) display theoretical virtues such as strength, unifying power, simplicity.

It's clear that traits (i) and (ii) of the logical notion of truth are not necessarily incompatible with semantic theorizing. Such an incompatibility was argued for by traditional deflationism on independent metaphysical grounds. The present volume features attempts to articulate a *compatibilist*

approach to the relationships between semantic applications and the logical notion of truth.

Disquotation, Compositionality, and Reflection. A hallmark of truth-conditional semantics is compositionality. It's useful to restrict our attention to sentences: compositionality requires that the truth-value of a sentence supervenes on the semantic values of its parts (e.g., a conjunction is true if and only if both conjuncts are true). Lavinia Picollo and Thomas Schindler, in *Is Deflationism Compatible with Compositional and Tarskian Truth Theories?*, investigate to what extent compositional principles can be compatible with the logical notion of truth. They propose two desiderata for theories of the logical notion of truth. The first, *Functionality*, enforces a mimimal adequacy requirement for the logical truth predicate. Since the truth predicate is—among other things—a generalizing tool, it has to satisfy the (uniform) T-schema for the class of sentences one wants to generalize over. The second desideratum, *Relative Insubstantiality*, prescribes that the principles characterizing the logical truth predicate should be derivable from the instances of the T-schema and some additional non-truth-theoretic principles. But what are these additional principles? And in what sense are they compatible with the logical truth predicate? Following recent developments in formal theories of truth, Picollo and Schindler consider *proof-theoretic reflection principles* and, in particular, Uniform Reflection Principles. In these recent developments, truth theorists investigate motivations and consequences of combining disquotational truth and proof-theoretic reflection. They show that Uniform Reflection can be used to derive compositional principles from disquotational ones (Horsten and Leigh, 2017; Fischer et al., 2017). Such developments are carefully surveyed in Horsten and Zicchetti's *Truth, Reflection, and Commitment*. Horsten and Zicchetti's contribution is not limited to the interplay between disquotation and reflection, but discusses also the mathematics of reflection principles and some of their recent philosophical applications.

Riki Heck's *Disquotationalism and Compositionality* and Johannes Stern's *Belief, Truth, and Ways of Believing* cast doubts on the compatibility of disquotational and semantic truth. One core tenet of classical disquotationalism is that *A* and '"A" is true' are fully cognitively equivalent. Their equivalence is stricter than material and necessary equivalence. Heck and Stern both argue that the strong equivalence required by disquotationalism overgenerates.[9] In particular, they both present cases to doubt this equivalence between *A* and '"A" is true' in non-extensional contexts.

Stern's criticism focuses on the status of the equivalence between *A* and '"A" is true' in belief contexts. He provides and discusses evidence supporting the claim that believing and believing-true need to be differentiated at the semantic level: for instance, someone may believe that Goldbach's conjecture is true, without believing Goldbach's conjecture,

because they lack relevant information concerning the representational status of 'Goldbach's conjecture'. However, Stern also provides a formal semantics that isolates a class of belief reports for which the equivalence envisaged by disquotationalists holds. When one is *aware of* appropriate facts concerning the syntactic representation of the belief under considerations, the disquotationalist's equivalence holds. We will come back to some aspects of Stern's formal model shortly.

Heck focuses on distinct kinds of overgeneration. The common theme of his criticism is the status of compositional principles in the disquotationalist's framework. First, compositional principles such as

(7) for all sentences *A*: '*A*' is not true if and only if '¬*A*' is true

should be taken to express—according to traditional disquotationalism—the trivial infinite conjunction of all the instances of the schema $\neg A \leftrightarrow \neg A$. However, and this is the first case of overgeneration, it's not clear why this infinite conjunction should not be expressed by

(8) for all sentences *A*: '¬*A* if and only if ¬*A*' is true.

(7) and (8) have very different logical properties. Therefore, they cannot both be taken to express—in the strong sense required by disquotationalism—the infinite conjunction of all instances of $\neg A \leftrightarrow \neg A$.

Heck's second overgeneration argument concerns directly the strategies employed by disquotationalists to recover compositionality from disquotation defended in Picollo and Schindler's contribution and surveyed by Horsten and Zicchetti. According to Heck, the same strategy that enables one to conclude (7) from the schema $\neg \mathrm{Tr}\ulcorner A \urcorner \leftrightarrow \mathrm{Tr}\ulcorner \neg A \urcorner$, also enables one to infer more dubious 'compositional' principles for intensional and hyperintensional connectives such as 'it's necessary that' and 'because'.

Compatibility and Contextualism. The purpose of the logical notion of truth consists in its expressive power. Above all, disquotationalists hold that the truth predicate is indispensable in blind ascriptions ('What Francis said on Sunday is true') and generalizations ('All theorems of Euclidean Geometry are true'). In turn, blind ascriptions and generalizations are necessary to the expression of agreement and disagreement (Field, 2008). Hartry Field forcefully argued that such an expressive role requires the truth predicate to be *transparent*: *A* and '"A" is true' should be intersubstitutable *salva veritate*. We have seen that transparency may be dubious in intensional and hyperintensional contexts, but disquotationalists maintain that it should at least be uncontroversial in extensional contexts.

In *The Expressive Power of Contextualist Truth*, Julien Murzi and Lorenzo Rossi put this last claim into question. They argue that

transparency is not required to perform blind ascriptions and gene-ralizations, and therefore to express agreement and disagreement. This leaves open whether weaker versions of the equivalence of *A* and '"A" is true' may be required for these purposes. In particular, one might still require the truth predicate to be *naïve*, in the sense that it satisfies the rule of inference 'from *A*, infer "'A' is true"', and its converse. Murzi and Rossi also argue that *naïveté* is not required by the logico-linguistic tasks considered by disquotationalists. They propose to employ a *contextualist* approach, and associated contextu-alist rules to model agreement and disagreement. In particular, they claim that the rules

(9) if '"A" is true' is inferred at context *α*, infer *A* at context *α*

(10) if ¬*A* is inferred at context *α*, infer 'it's not the case that
     "A" is true' at *α*

suffice to adequately model the cases of agreement and disagreement that arise in the disquotationalist literature.

Murzi and Rossi's proposal can be seen as an alternative way of achieving some form of compatibility of the logical notion of truth with semantic truth. Contextualist approaches to truth and paradox draw much of their motivation from ideas in philosophy of language and linguistics. A central task for the contextualist is to determine con-texts in which specific truth ascriptions have definite semantic values. In particular, according to the contextualist, there are contexts (such as paradoxical ones) in which a given truth ascription is semantically defective because it fails to express a proposition, but it may express a proposition in other contexts. In such an alternative context the truth ascription will have a definite semantic value.

By arguing that this semantically loaded notion of truth can be recon-ciled with some of the tasks intrinsic to the disquotationalist's truth pred-icate, Murzi and Rossi create a promising bridge between the two concepts of truth analysed in the first part of the volume.

1.2 Unification: Formal Semantics

Whatever principles of truth one chooses, a core tenet of the Unified Approach investigated in this volume is that truth should naturally interact with modal and doxastic notions. This interaction should then be modelled by a satisfactory formal semantics. These desiderata immediately raise problems. The standard semantic framework for modal and doxastic notions is *possible worlds semantics*. The prominence of possible worlds semantics can be easily explained by some of its theoretical virtues: it's simple and widely adaptable to various domains, it fits strong pre-theoretic intuitions concerning alternative metaphysical and doxastic scenarios, and

it has proven to be successful in modelling phenomena in linguistics, philosophy, and computer science. Possible worlds semantics is usually applied to languages in which modal and doxastic notions are formalized as sentential operators. As such, they apply to formulae that have a finite (well-founded) structure.[10] The satisfaction of a formula φ at a world is defined there by a straightforward induction on φ's finite syntactic structure. However, this strategy does not extend so easily to languages containing a truth predicate. Due to the possibility of self-referential constructions, objects to which the truth predicate applies may have an infinite syntactic structure. Therefore, one requires more involved strategies. The inductive strategy of fixed-point models in the style of Kripke (1975) is one example; another is the class of models obtained by revision-theoretic strategies from Gupta and Belnap (1993).

It's however possible to combine semantic constructions for (self-referential) truth predicates and possible worlds semantics. The basic idea is simple. For definiteness, let's focus on the interaction of truth and necessity predicates. One can think of a *frame F* as given by a collection of standard interpretations of our basic language $\mathcal{L}$ (without truth nor necessity)—the collection of *worlds*—together with some binary accessibility relation R on this collection. The interpretations of the truth and necessity predicates are given by a suitable function f that assigns (suitable) sets of sentences to each world in the frame. Given a world w, the intended role played by the evaluation function is to provide both the extension $f(w)$ of the truth predicate at w, and the extension

$$\bigcap_{wRv} f(v)$$

of the necessity predicate at w, that is, the intersection of the extension of the truth predicate at all worlds accessible from w.

To construct suitable evaluation functions, one can apply the strategies employed in the resolution of the liar paradox. For instance, as shown in Halbach and Welch (2009), one can generalize Kripke's fixed-point semantics and define a suitable evaluation function as the fixed point of a positive inductive definition that is performed simultaneously at each world.[11] This procedure will yield suitable evaluations for arbitrary frames. If one does not define the evaluation function via some positive inductive construction, one is not guaranteed that suitable evaluations will be found for arbitrary frames. However, as shown in Halbach et al. (2003), converse well-founded frames always admit evaluations.

In the framework just outlined, many interesting philosophical principles can be validated in full generality. For instance, models based on generalized fixed points will satisfy the factivity of necessity:

(11) $\forall x(\text{Sentence}(x) \land \text{Nec}(x) \rightarrow \text{True}(x))$.

Claims such as (11) deal with sentences—or whatever objects of truth and modality one plausibly assumes instead of sentences. They are *de dicto* assertions. Even in expressive frameworks such as the semantics just sketched, *de re* ascriptions are more difficult to treat. In the models considered by Halbach and Welch (2009), Stern (2015), and Stern's contribution to this volume, one has syntactic names for all objects in the domain of discourse. In those happy circumstances, a *de re* version of factivity can be readily formulated because the formal syntax is sufficiently rich to support quantifying into modal contexts:

(12) $\forall x, y(\text{Sentence}(x(y)) \wedge \text{Nec}(x(y)) \rightarrow \text{True}(x(y)))$.

(12) expresses that if the predicate x is necessary of y, then it is also true of y.

   The assumption that our language can name all objects, especially when dealing with general claims involving logical and metaphysical necessity, is clearly too strong: our language does not contain names for all possible objects. Volker Halbach, in *The Fourth Grade of Modal Involvement*, investigates languages in which *de re* ascription can be formalized in full generality. The fundamental idea is to generalize Tarski's (1956) treatment of satisfaction to modal predicates. *De re* factivity can now be captured by means of *binary* predicates for necessity and truth:

(13) $\forall x, y(\text{Formula}(x) \wedge \text{Nec}(x, y) \rightarrow \text{True}(x, y))$.

(13) expresses that if the object y possesses the property expressed by x necessarily, then x is also true of y. Halbach outlines different alternatives for a possible worlds semantics for *de re* necessity, and discusses some metaphysical applications of the framework. In (13), the object y may stand for an arbitrary (finite) sequence of objects in the domain of discourse. One crucial aspect of the semantics sketched by Halbach is the strict link between object-linguistic sequences figuring in *de re* ascriptions and metalinguistic sequences of variable assignments. Halbach provides ingenious examples to show that, once the full power of the Fourth Grade of Modal Involvement is available, mathematical assumptions of the metatheory are reflected in one's theorizing in the object-language: if one carelessly adopts the standard metatheory of quantified modal logic, strong theses such as actualism may follow.

   In *Belief, Truth, and Ways of Believing*, Johannes Stern considers the interaction of truth and belief. We have already seen how Stern's contribution can be seen as a reaction to some shortcomings of the disquotationalist's position. However, his work also contains a detailed formal semantics for a language with a truth predicate, an awareness predicate, and a belief operator. Stern's model is based on the generalized fixed-point strategy outlined above, but extends it with a suitable

interpretation of the belief operator. The main ideas are to extend the notion of evaluation and the associated notion of satisfaction from worlds to *pairs of worlds* (w, v), and to parametrize the accessibility relation to the agents' belief representation. The first component of the pair (w, v) is the one that provides the location in the possible worlds structure at which the basic vocabulary, the truth predicate, and the awareness predicate are evaluated; the second component of the pair fixes the location of the agent's awareness set at a world, which is the set to which the accessibility relation is parametrized. For instance, to evaluate a sentence of the form

(14) *S* believes that *S* believes that 'snow is white' is true

at (w, v), one looks at all of *S*'s doxastic alternatives z with respect to w and at *S*'s awareness set at v to evaluate '*S* believes that "snow is white" is true'. However, to evaluate '"snow is white" is true' at all doxastic alternatives u to z, one will look at the pair (u, v): the location v of the awareness set is kept fixed. In Stern's analysis, this yields a correct way of capturing the semantic resources that the agent has at her disposal in evaluating truth ascriptions: someone may in fact believe a proposition p, but may not believe the truth ascription $\text{Tr}\ulcorner \varphi_p \urcorner$, because she may not be aware that φ_p expresses the proposition p.

The strategies employed to provide a semantics for expressive languages allowing for self-reference can also shed light on some puzzles of rationality. In her contribution *Indeterminate Truth and Credences* Catrin Campbell-Moore focuses on so-called self-undermining credences. A credence is called self-undermining if, in case a rational agent were to adopt it, they would immediately be compelled to revise their attitude. It turns out that in certain scenarios all *precise* credences will be self-undermining, just like all assignments of a classical truth-value to the liar sentence would result in the opposite value. But one may then ask which attitude a rational agent ought to adopt, and it is at this point that the strategies for addressing the semantic and intensional paradoxes considered above come into play.

After developing a revision jump in the style of Gupta and Belnap (1993) for definite assignments of truth values, Campbell-Moore defines a revision jump for sets of precise credences to model the relevant cases of self-undermining credences: the revision process in such cases does not reach a fixed point. To find non-self-undermining credences, Campbell-Moore then moves to imprecise credences (sets of precise credal functions) and develops a novel, generalized fixed-point semantics based on Kripke's supervaluational jump. Similarly to what happens to the liar sentence in Kripkean fixed points, self-undermining credences are indeterminate, in the sense that no definite credence can be associated with them in this framework. Campbell-Moore's semantics for

self-undermining credences provides a rich an interesting model of belief. The framework also stresses that self-reference is a phenomenon that needs to be taken seriously, as it arises in various and sometimes unexpected contexts.

1.3 Unification and Higher-Order Resources

As argued above, the Unified Approach draws from the analysis of quantification in natural language. It appears that in natural language there is only one fundamental kind of quantification, that is quantification into singular-term position. We have seen that it is possible to quantify over nominalized properties and relations in suitable frameworks inspired by the Unified Approach, such as Halbach's framework for *de re* necessity. Some authors forcefully argued that first-order quantification is not sufficient to theorize at the level of generality required by some areas of philosophy. Instead one should adopt a new, irreducible form of quantification over higher-order entities (Williamson, 2003).

Advocates of the higher-order perspective typically motivate their position by noting that a faithful semantics for absolutely general theses such as 'Everything is self-identical' cannot be given in standard (first-order) set theory. Instead, one should provide models that aren't sets, but *sui generis* entities of an irreducibly higher-order type. This irreducibility is typically motivated by higher-order versions of Cantor's theorem, stating that such higher-order entities are also of a different size than first-order ones.

A reaction from the point of view of the Unified Approach may focus on the inherent expressive limitations of the higher-order perspective. Entities are syntactically distinguished in a hierarchy of types: such types are metatheoretic objects that, despite their essential role in the categorization of the type-theorists' ontology, are not part of the domain of higher-order quantifiers. Moreover, the type-theorist would hope that such commitment to types could be reduced to the minimum: namely, that one could resort to a coherent rationale to endorse only arbitrarily many finite types, as in Simple Type Theory. However, there are arguments suggesting that such a rationale may be difficult to find. Linnebo and Rayo (2012) argue that the type-theorist is bound to countenance proper class-many types.

James Studd's *Infinite Types and the Principle of Union* reacts to Linnebo and Rayo's argument. Their argument rests crucially on the claim that the type-theorist does not have at her disposal a principled way to stop ascending the hierarchy of types at any limit ordinal. Studd carefully examines this claim, the Principle of Union, and finds it wanting. In particular, he argues that the union language, say $\cup_{n \in \omega} \mathcal{L}_n$ with n a finite type, has a special semantic status. One could provide a semantics for each of its sublanguages, but this is not yet

sufficient ground to countenance a language of type $\omega + 1$ without assuming the Principle of Union itself. This is no more justified than granting, without an Axiom of Infinity, a *set* of all finite sets on the basis of the acceptance of all finite sets.

We have seen that higher-order resources, if not strictly regimented in a hierarchy of types, allow for great expressive power that lead to paradoxical phenomena akin to the semantic and modal paradoxes. Andrew Bacon's *Opacity and Paradox* investigates the so-called Prior's Paradox for Thought, which is one instance of the paradoxes of indirect discourse already encountered in this introduction. Prior's Paradox is actually a theorem of propositionally quantified intensional logic and involves an arbitrary sentential operator, which can be interpreted as 'Mary thought at time t that'. Prior's Theorem then states that, if Mary thought at time t that all she thought at time t was false, then Mary thought a true and a false thing at t. The puzzling feature of this theorem comes from the fact that it's plausible to think that Mary thought only one thing at t, namely that all she thought at time t was false. But in that case, she couldn't have thought a true and a false thing.

Bacon proposes a solution to Prior's paradox based on the restriction of Universal Instantiation in opaque contexts such as 'it is thought by Mary at t that'. The idea is familiar from the cluster of Frege's puzzles and the puzzles of belief discussed in Stern's contribution: given two semantically equivalent (possibly identical) expressions, this equivalence may fail to be reflected in opaque contexts. He also provides models that invalidate Prior's Theorem (and Universal Instantiation), but preserve many other desirable principles of propositional quantification. Interestingly, such a framework enables him to suggest a resolution of the liar paradox akin to classical gap-theories proposed by Feferman (1991) and Maudlin (2004). Bacon defines a truth predicate on the basis of the (opaque) context 'means that'. The liar sentence *says that* it is not true, but *there's nothing it says*. So, the liar sentence conveys information that can be acted upon, and even known. However, due to the opacity phenomenon, such information does not pick out semantic content: there's no proposition corresponding to the liar.

Acknowledgments

Notes

1. Throughout the Introduction we adopt a wide understanding of 'first-order quantification' that, e.g., includes the use of generalized quantifiers. The

relevant contrast is between (first-order) quantification, which, assuming some form of Montague Grammar, binds argument positions of type *e*, and higher-order quantification binding argument positions of higher types. See Heim and Kratzer (1998) or Gamut (1991) for an introduction to Montague Grammar.

2. This may be an oversimplification. Perhaps one needs to accommodate a rich ontology of various attitudinal objects à la Moltmann (2017a, b).
3. That is, unless these expressions are of flexible type.
4. This does not imply that modality and modal notions need to be generally conceived to be of the same logical/grammatical category as truth. It is perfectly acceptable and perhaps desirable to, following Kratzer (1981), analyse modality, i.e., the modal aspect of sentences of natural language via a modal operator but truth as a predicate. The point is that in sentences like 'Everything necessary is true', 'is necessary' needs to be conceived of as a predicate if truth is, i.e., 'is necessary' needs to be construed as a phrase of type $\langle e, t \rangle$.
5. See Stern (2015) for an overview of the various paradoxes.
6. See specifically Stern and Fischer (2015) for a discussion of this point.
7. The paradoxes of indirect discourse have been re-discovered by Brandenburger and Keisler (2006) and are known under the label the Brandenburger-Keisler paradox in the literature on epistemic game theory.
8. How to understand this equivalence is a non-trivial matter. We will come back to this point below.
9. Issues of overgeneration have been raised famously by Gupta (1993) and, more recently, by Nicolai (2020).
10. Formulae are generally modelled after finite, well-founded trees or sequences that unravel their syntactic structure.
11. For the frame ($\{w\}$, $\langle w, w \rangle$), a suitable evaluation *f* consists simply in a function that assigns a Kripke-fixed point to *w* to both the truth of necessity predicate. The interpretations of the two predicates will diverge in more complex frames, but it will always yield fixed points.

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.

Beall, J. C. (2009). *Spandrels of Truth*. Oxford 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.

Bull, R. A. (1969). Modal logic with propositional quantification. *The Journal of Symbolic Logic*, 34(2): 257–263.

Cobreros, P., Égré, P., Ripley, D., and van Rooij, R. (2013). Reaching transparent truth. *Mind*, 122(488): 841–866.

Cross, C. B. (2001). A theorem concerning syntactical treatments of nonidealized belief. *Synthese*, 129: 335–341.

Feferman, S. (1991). Reflecting on incompleteness. *Journal of Symbolic Logic*, 56: 1–49.

Field, H. (2008). *Saving Truth from Paradox*. Oxford University Press.

Fine, K. (1970). Propositional quantifiers in modal logic. *Theoria*, 36: 336–346.

Fischer, M., Nicolai, C., and Horsten, L. (2017). Iterated reflection over full disquotational truth. *Journal of Logic and Computation*, 27(8): 2631–2651.

Gamut, L. (1991). *Logic, Language, and Meaning, Volume 2: Intensional Logic and Intensional Grammar*. The University of Chicago Press.

Grover, D. (1992). *A Prosentential Theory of Truth*. Princeton University Press.

Grover, D., Camp, J., and Belnap, N. (1975). A prosentential theory of truth. *Philosophical Studies*, 27: 73–125.

Gupta, A. (1993). A critique of deflationism. *Philosophical Topics*, 21(1): 57–81.

Gupta, A. and Belnap, N. (1993). *The Revision Theory of Truth*. MIT Press.

Halbach, V. (2006). How not to state T-sentences. *Analysis*, 66(4): 276–280.

Halbach, V. (2008). On a side effect of solving Fitch's paradox by typing knowledge. *Analysis*, 68(298): 114–120.

Halbach, V., Leitgeb, H., and Welch, P. (2003). Possible-worlds semantics for modal notions conceived as predicates. *Journal of Philosophical Logic*, 32 (2): 179–223.

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 (469): 71–100.

Heim, I. and Kratzer, A. (1998). *Semantics in Generative Grammar*. Blackwell Publishing.

Horsten, L. and Leigh, G. E. (2017). Truth is simple. *Mind*, 126(501): 195–232.

Horsten, L. and Leitgeb, H. (2001). No future. *Journal of Philosophical Logic*, 30(3): 259–265.

Kratzer, A. (1981). The notional category of modality. In Eikmeyer, H.-J. and Rieser, H., editors, *Words, Worlds, and Contexts: New Approaches to World Semantics*, pages 38–74. de Gruyter.

Kripke, S. (1975). Outline of a theory of truth. *Journal of Philosophy*, 72: 690–712.

Linnebo, O. and Rayo, A. (2012). Hierarchies ontological and ideological. *Mind*, 121(482): 269–308.

Maudlin, T. (2004). *Truth and Paradox: Solving the Riddles*. Oxford University Press.

Moltmann, F. (2017a). A truthmaker theory for modals. *Philosophical Issues*, forthcoming.

Moltmann, F. (2017b). *Truth Predicates, Truth Bearer, and their Variants*. http://friederike-moltmann.com/uploads/TruthPredicates-July-publ-2017.pdf.

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

Mulligan, K. (2010). The truth predicate vs. the truth connective. On taking connectives seriously. *Dialectica*, 64(4): 565–584.

Myhill, J. (1960). Some remarks on the notion of proof. *Journal of Philosophy*, 57(14): 461–471.

Nicolai, C. (2020). Fix, express, quantify. Disquotation after its logic. *Mind*. Online version.

Priest, G. (2006). *In Contradiction: A Study of the Transconsistent*. Oxford University Press.

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.

Quine, W. V. O. (1970). *Philosophy of Logic*. Harvard University Press.

Ramsey, F. P. (1927). Facts and propositions. *Proceedings of the Aristotelian Society*, 7: 153–170.

Ramsey, F. P. (1929). The nature of truth. In Rescher, N. and Majer, U., editor, *On Truth: Original Manuscript Materials (1927–1929) from the Ramsey Collection at the University of Pittsburgh*. Kluwer Academic Publishers. Collection published in 1991.

Ripley, D. (2012). Conservatively extending classical logic with transparent truth. *Review of Symbolic Logic*, 5(2): 354–378.

Stern, J. (2015). *Toward Predicate Approaches to Modality*, volume 44. Springer.

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

Tarski, A. (1956). Der Wahrheitsbegriff in den formalisierten Sprachen. In *Logic, Semantics, Metamathematics*, pages 152–278. Clarendon Press.

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

Williamson, T. (2003). Everything. *Philosophical Perspectives*, 17(1): 415–465.