

A LANDSCAPE OF LOGICS BEYOND THE DEDUCTION THEOREM (AND MOORE’S PARADOX)

BAS C. VAN FRAASSEN

Princeton University and San Francisco State University, USA
 fraassen@princeton.edu

Abstract. Philosophical issues often turn into logic. That is certainly true of Moore’s Paradox, which tends to appear and reappear in many philosophical contexts. There is no doubt that its study belongs to pragmatics rather than semantics or syntax. But it is also true that issues in pragmatics can often be studied fruitfully by attending to their projection, so to speak, onto the levels of semantics or syntax — just in the way that problems in spherical geometry are often illuminated by the study of projections onto a plane. To begin I will describe a potentially vast landscape of logics of a certain form, with some illustrations of how they appear naturally in response to some problems in philosophical logic. Then I will turn Moore’s Paradox into logic, within that landscape, and show how far it can be illuminated therein.

Keywords: supervaluation • deduction theorem • logic of belief • Moore’s Paradox

RECEIVED: 24/11/2021

ACCEPTED: 22/02/2022

1. What is classical logic?

When I first taught classical sentential logic it was, as is now usual, in natural deduction format, and was scarcely aware how large a change that marked from its origin. The original classical sentential logic (henceforth **OCSL**), as found in Whitehead and Russell (1910) for example, was essentially what we now call a Hilbert style system: a system of axioms, with a single derivation rule, Modus Ponens.¹

How did the profession arrive at the current teaching form, classical sentential logic in natural deduction formulation (henceforth **NDFSL**)? And in what sense are **OCSL** and **NDFSL** different formulations of the same logic — if, indeed, they are?

2. The Watershed: Herbrand 1930

The change certainly derives from the innovations introduced by Gentzen (1934-5) and Jaskowski (1934). But those innovations had their logical origin in Herbrand’s Deduction Theorem (Herbrand 1930):



(*) $X, A \vdash B$ iff $X \vdash A \supset B$.

This was proved for **OCSL**. What it implied for logical practice was that the

Rule of Conditional Proof: If B is derivable from premises X together with premise A then $(A \supset B)$ is derivable from X

is *admissible* for **OCSL**. That is to say: adding the Rule of Conditional Proof to **OCSL** does not change the set of direct inferences. To be precise, for the augmented system of **OCSL** with Conditional Proof, the following holds:

$\vdash A$ in **OCSL** iff $\vdash A$ in the augmented system;
 $X \vdash A$ in **OCSL** iff $X \vdash A$ in the augmented system.

So, if we take these two equivalences as the criteria for being the same logic, then **OCSL** and **NDFS** are the same logic.

As you will no doubt conclude from the emphasis in my text, I will be exploring how they may not be the same.

I call Modus Ponens a *direct inference rule* for it sanctions only moves from given premises to a given conclusion. Conditional Proof is not a direct inference rule, it is an *inference to inference rule*. Once added the logic is no longer a Hilbert style system, it is something else.

Other inference to inference rules were quickly shown to be admissible, for example Proof by Cases (Disjunctive Syllogism), and the route to the current Natural Deduction Formulation was now open.

All of this was about the old familiar sentential logic. A quick look at the proof of the Deduction Theorem shows that it hinged on the system having, besides its axioms, just that one direct inference rule of Modus Ponens. The question is then: under what conditions can the Deduction Theorem be proved in general?

3. Applied Logic

In philosophical logic, and even more in philosophy of science, our concern is with languages in which the terms and sentences have meaning or reference, that is, with interpreted or semi-interpreted languages. Classical sentential logic may be sound for such a language, but the meaning relations among the terms and sentences yield other valid inferences, to be captured within the logic of that language.

Rudolf Carnap (1952), concerned with the character of any such applied logic, introduced the notion of *meaning postulates*: additional axioms, not logically true, but true *ex vi terminorum*, as the Medievals would have said.

Let us call such a logic an *axiom-extended* logic. If non-logical axioms are added to **OCSL**, the Deduction Theorem continues to hold. More generally, if the Deduction Theorem holds for a logic then the Theorem will also hold for all its axiom extensions.

But what if there are, in such a language, arguments not logically valid but valid *ex vi terminorum* and a corresponding rule is added? Let us call that a rule-extension and the resulting logic a *rule-extended* logic. If the Deduction Theorem held in the original, will it still hold in the rule-extension?

As we know, the addition of rules in this way can trivialize a logic. Arthur Prior (1960) gave the example of a non-logical connective, *tonk*, with the introduction and elimination rules

$$\begin{array}{l} A \vdash (A \text{ tonk } B) \\ (A \text{ tonk } B) \vdash B \end{array}$$

So some rules cannot be added on pain of triviality. On the other hand, if a rule-extension of a logic is not trivial or defective in some other way, is it perhaps always equivalent to an axiom-extension? In that case the acceptable rule-extensions would all be such as to present no obstacle to the continued validity of the Deduction Theorem. It would also, of course, trivialize the distinction between axiom-extension and rule-extension.

4. An example of rules in conflict with the Deduction Theorem

Just as Prior did, let me introduce a non-logical connective, **T**. This one is unary, and is to be read as “It is true that”. However, I propose to have this in a language which is not bivalent: it has truth-values 1 and 0 but also some other ones. Only 1 is designated: all the others are ways of not being true, and a sentence is valid if and only if it must always have value 1. Negation acts in a classical way as long as only 1 and 0 are involved. That is, *A* has value 0 only if $\sim A$ has value 1, and *A* has value 1 only if $\sim A$ has value 0. Finally, a disjunction has value 1 if and only if both disjuncts have value 1. Since a sentence can have a different value from 1 or 0, it is clear that $(TA \vee T\sim A)$ is not a valid sentence.

Could there be such a language as I have just described? William and Martha Kneale, in their great history of logic, offered a proof that there cannot be (Kneale 1962: 47). Let us examine this.

What I have described implies very clearly that the following arguments are valid in this language:

$$\begin{array}{l} A, \text{ therefore } TA \\ TA, \text{ therefore } A \end{array}$$

So the logic of this language should have the direct inference rules

$$A \vdash \mathbf{T}A$$

$$\mathbf{T}A \vdash A$$

Looking at the Kneales' text, it is clear that they assumed that the Deduction Theorem must hold. In that case, the rule-extension in question is equivalent to an axiom-extension: those inference rules yield the conditionals of forms $(E \supset \mathbf{T}E)$ and $(\mathbf{T}E \supset E)$ as theorems. With that assumed, the Kneales' proof is correct:

1. $A \supset \mathbf{T}A$ theorem
2. $\sim \mathbf{T}A \supset \sim A$ [1]
3. $\sim A \supset \mathbf{T}(\sim A)$ theorem
4. $\sim \mathbf{T}A \supset \mathbf{T}(\sim A)$ [2, 3]
5. $\mathbf{T}A \vee \mathbf{T}(\sim A)$ [4], definition of " \vee "

and so the Principle of Bivalence has been proved a priori!

But I submit that the sort of language I described, which is not bivalent, can (and does) exist.²

Our challenge, then, is to provide a general account of rule-extended logics, and of the languages of which they are the logics: how to understand them, and how to characterize the valid argument relation in such languages.

5. A general semantic analysis for rule-extended logics

I will examine a Hilbert style system, with a corresponding language, and a rule-extension of that system. Then I will provide a semantic analysis of the language for which that rule-extension is sound and complete. I will outline this as a general procedure, for the construction of a language in which the rule-extension is sound, when starting with a given Hilbert style system.

5.1. Starting with a basic system **S**

System **S** is an axiom-extension of **OCSL**, and is sound and complete for a corresponding language **LS**.

The language **LS** has a class of sentences, and a class of admissible valuations (to be referred to as the *classical valuations* of **LS**). These valuations assign one of the values T, F to each sentence, and are truth functional, by the classical truth-table rules, for \supset and \sim (and other connectives defined in terms of these in the classical manner).

We shall call a set of sentences in the syntax of **LS** *classically satisfiable* if and only if there is some classical valuation which assigns T to all the members of that set. The valid consequence relation in this language is thus defined as follows:

$X \models A$ in **LS** if and only if all classical valuations which satisfy X also satisfy A .

5.2. Introducing non-logical rules and the rule-extension **ExL**

Imagine that at this point we gain a greater insight in the language that we had formalized as **LS**, and that we realize that certain non-logical direct inference rules **R** must be recognized as valid.

The classical valuations do not respect the non-logical rules in **R**. Let **ExL** be the rule-extension of **S** formed by adding rules **R**.

Since **ExL** will not be a Hilbert style system, it should be noted that the definition of derivability for the rule-extension takes the same form as for **S**:

Definition. A is *derivable* from X in **ExL** ($X \vdash A$ in **ExL**) if and only if there is a finite sequence of sentences, each of which is either a member of X or an axiom of **S** or follows from preceding members either by Modus Ponens or by one of the rules in **R**.

Definition. An **ExL** *theory* is a set of sentences in the syntax of **LS** which is closed under derivability in **ExL**.

It follows from the definitions that $X \vdash A$ in **ExL** if and only if A belongs to all **ExL** theories that contain X .

How shall we approach the semantic analysis of this system, that is, how shall we conceive of a language **LS*** with the syntax of **LS** in which

- (a) the theorems of system **S** are still valid sentences,
- (b) the inference rule of **S**, that is to say Modus Ponens, is still truth-preserving,
- (c) the inference rules in **R** are also truth-preserving?

The answer to this question is relatively straightforward. The ‘possible world’ descriptions — the consistent satisfiable sets of sentences — must be classically satisfiable **ExL** theories. These are of course the sets of sentences that include all theorems of **S**, closed under Modus Ponens, and also closed under all the rules in **R**, and are satisfied by some classical valuation.

Equivalently, the admissible valuations of **LS*** are valuations that satisfy the classically satisfiable **ExL** theories.

What can these be?

5.3. Enter supervvaluations

We define the class of admissible valuations of language \mathbf{LS}^* as follows (they are *supervvaluations* of \mathbf{LS}).

Definition. φ is an *admissible valuation* of language \mathbf{LS}^* iff there is a classically satisfiable theory X of \mathbf{ExL} such that, for all statements A :

- $\varphi(A) = \text{T}$ iff all classical admissible valuations which satisfy X , assign T to A ;
 $\varphi(A) = \text{F}$ iff all classical admissible valuations which satisfy X , assign F to A .

In all other cases, $\varphi(A)$ is undefined.

These functions are thus in general partial functions on the set of sentences. When a sentence is not assigned a truth-value by such an admissible valuation of \mathbf{LS}^* , we speak of a ‘truth-value gap’. The admissible valuations of \mathbf{LS}^* can be equivalently characterized as follows.

Definition.

- (a) $\text{CLAS}(X)$ = the set of classical admissible valuations which satisfy X ;
 (b) φ is an *admissible valuation* iff there is a classically satisfiable \mathbf{ExL} theory X such that $\varphi = \bigcap \text{CLAS}(X)$.

We may just remark here that this entails a distinction between the *Principle of Bivalence* (that every sentence is true or false) and the *Principle of Excluded Middle* (that, for every sentence A , the sentence $(A \vee \sim A)$ is logically true. For all classical tautologies are still true on all supervvaluations, and all classical direct inferences, from premises to a conclusion, also remain valid.

Definition. $X \models A$ in \mathbf{LS}^* if and only if all admissible valuations of \mathbf{LS}^* which satisfy X also satisfy A .

5.4. Soundness and Completeness Theorem

We take it as given that system \mathbf{S} is sound and complete for language \mathbf{LS} . Assume for a moment that there are admissible valuations for the language \mathbf{LS}^* . (That we have defined them does not guarantee their existence: since we have not put constraints of the set \mathbf{R} of non-logical rules, there might not be any classically satisfiable \mathbf{ExL} theories).

Theorem. If $X \models A$ in \mathbf{LS} then $X \models A$ in \mathbf{LS}^* .

This is clear from the definitions, since a counterexample to an argument which is a supervvaluation of \mathbf{LS} would require there to be a classical valuation that is a counterexample to that argument.

Theorem. $X \models A$ in \mathbf{LS}^* if and only if $X \vdash A$ in \mathbf{ExL} .

By the definition of admissible valuation of \mathbf{LS}^* it follows that $X \models A$ exactly if A belongs to all classically satisfiable theories of \mathbf{ExL} that include X . There may well be \mathbf{ExL} theories that are not classically satisfiable. But the theorem follows from:

Lemma. If A belongs to all classically satisfiable \mathbf{ExL} theories which include X then A belongs to all \mathbf{ExL} theories which include X .

For suppose the antecedent, and that Y is an \mathbf{ExL} theory which contains X but is not classically satisfiable. Then by the soundness and completeness of \mathbf{S} for \mathbf{LS} and the fact that all axioms and rules of \mathbf{S} belong to \mathbf{ExL} , Y is not a consistent \mathbf{S} theory. Since \mathbf{S} is an axiom extension of \mathbf{OCSL} it follows that Y contains all sentences. Hence it also contains A .

The upshot? Soundness and completeness for a rule-extension are immediate, in all cases for which our assumptions hold.

So where does the work come in? The assumption we introduced at the beginning was that \mathbf{LS}^* actually has admissible valuations. If it does not, soundness and completeness are trivial.

To establish *non-triviality*, it is required to show that, in view of what the rules in \mathbf{R} are, the language \mathbf{LS}^* actually does have admissible valuations. And more of course: in all cases of interest it is required to go beyond this minimal requirement, and to show that this language has in fact a large, rich array of admissible valuations, so that all that we consider relevantly significant can be expressed in this language.

5.5. General characteristics of rule-extensions and failure of inference-to-inference rules

What remains, of familiar inference to inference rules?

To begin there are the very basic ones that make no reference to details in the syntax, they are:

The Structural Rules:

if A is in X then $X \models A$

if $X \models A$ then $X, B \models A$

if $X \models A$ and $Y \models B$ for all B in X then $Y \models A$

Any other inference to inference rule can be invalidated in some rule-extended logic (van Fraassen 1971, Ch. V, 3c.).

At the end I will take up some counterexamples to familiar natural deduction rules, the ones — like Conditional Proof — that rely on previous inferences. By that

time we will have a concrete example of a rule-extended logic, based on considerations about Moore's paradox, to give us live examples.

6. Moore's Paradox and Epistemic Logic

My beliefs are often false, but I have difficulty acknowledging this. And there are many truths that I do not happen to believe — but that too is difficult to express with precision and rigor. G. E. Moore pointed this out (Moore 1942a: 543 and 1942b: 204), and the point was taken up by Wittgenstein. As an example let us take:

- (a) It is now raining in Peking, and I do not believe that it is now raining in Peking.
- (b) It is not raining in Peking now, but I do not believe that it is not raining in Peking now.

It is clear that if either of these assertions was one I could make truly and sincerely, it would show a curious sort of incoherence in my beliefs. For asserted truly and sincerely it would reveal that there was something that I both did and did not believe.

Yet neither of these two statements expresses something logically false. In fact, since at present I neither believe nor disbelieve that it is now raining in Peking it follows that either (a) is true or (b) is true.

Let me use small letters like "*p*" now to stand for statements and introduce the unary connective *B* to stand for "I believe that". Then the above shows that the form (*p* & ~*Bp*) is not the form of a logical falsehood, although it is not something an agent could ever truly assert.

Things look even worse for the form (~*p* & *Bp*). This appears to express belief in an acknowledged falsehood, and so again, if asserted truly and sincerely would reveal incoherence in my beliefs. Yet again the form is not the form of a logical falsehood, since I must acknowledge, that unfortunately, as I sometimes find out in retrospect, I do tend to have false beliefs.

The most obvious response from the side of logic is that this is a matter which belongs neither to proof theory nor to semantics, but to pragmatics. For the offending sentences are in first-person, and cease to be problematic if "I" is replaced by "she".

That is correct, and the subject of pragmatic tautologies and absurdities is fascinating in itself. But it is still possible, and may be fruitful, to attempt to formalize this language and to investigate, at the shallow level of semantics and syntax, the conditions under which a believer will never fall prey to Moore's paradox.

That is what I shall do now but, and this is the relevance to all that went before, I will show that there is just under the surface a rule-extended logic. In fact, it is the logic that governs the agent's own reasoning as she manages her own beliefs.

6.1. Attempts to uncover a hidden contradiction

In the literature, starting with Hintikka (1962: 16–31, 88–89, 123–125), we can see suggestions that Moore’s paradox is to be handled in the logic of belief by strengthening the axioms. The idea seems to be this:

When I believe something, whatever it is, I believe it to be true. So, whatever I believe I believe to be true. Therefore I do not believe that $(Bp \ \& \ \sim p)$ because I believe its opposite, namely $(Bp \supset p)$. This would certainly warrant the acceptance of the following axiom in the logic of belief:

$$(*) \vdash B(Bp \supset p)$$

Jaakko Hintikka, Richmond Thomason, and Charles Cross all considered that this axiom should be part of the logic of belief, referring to Moore’s paradox for its inspiration (see van Fraassen 2011 for details). But then Thomason showed that in a sufficiently rich context, for a sufficiently self-aware agent, this leads to a contradiction via a Goedel type argument (Thomason 2011; van Fraassen 2011).³

I submit that the clue for us lies here in my phrase “a sufficiently self-aware agent”, and want to suggest that sufficient self-awareness would lead an agent’s discourse not to be subject to suggested axiom (*), but to another somewhat similar looking one.

6.2. A self-transparent believer

If a robot were constructed to simulate rational belief and were programmed only to avoid logical falsehoods and unsatisfiable theories, it would have no qualms about entering into instances of Moore’s paradox. To have the insight which avoids the paradox it would have to record what its beliefs are (so far) and then become aware of the discordance.

Accordingly I suggest that we focus on believers who, though perhaps wrong about many things, *are right about what beliefs they have*. With the notation introduced above, that would require the logical principles

$$1] \text{ If } Bp \text{ then } BBp$$

$$2] \text{ If } BBp \text{ then } Bp$$

to hold. I will call such believers *self-transparent*.⁴

To begin I will discuss this in the context of normal modal logic. I imagine possible worlds, each with a designated agent. There are two ways of looking at such a possible world model. For us, looking from the outside, the sentence Bp is true in world w exactly if the designated agent in w believes that p . This agent would be ‘saying in his heart’ “I believe that p ”. So for us, Bp is a sentence in third-person language, while the agent assents to the corresponding first-person sentence.

6.3. The normal modal logic KD4C4

The family of normal modal logics is so well studied that we have a complete taxonomy with a corresponding systematic nomenclature. The logic of the self-transparent believer is an extension of the basic system **K**, with the addition of:

the Deontic principle: If Bp then $\sim B\sim p$

the **S4** principle: If Bp then BBp

the converse **S4** principle: If BBp then Bp

Hence it is designated **KD4C4**. In the standard semantic analysis of normal modal logics the model structures have general form: a set of possible worlds with an access relation **R** among them. The assignment of truth-values is the classical bivalent one for \sim and \supset , while Bp is true in a world w exactly if p is true in all worlds bearing **R** to w .⁵

For each of the specific logical principles there is a corresponding condition on **R**. For the deontic principle it is that **R** is *serial*: for each world w there is at least one world that bears **R** to w , and for the **S4** principle it is that **R** is *transitive*. What is of interest to us here is the condition that corresponds to the converse **S4** principle. It has various names in the literature, but since it is implied by reflexivity I give it a suitably mnemonic name:

Definition. The access relation **R** is *weakly reflexive* if and only if for any worlds x , y , if xRy then there is a world z such that xRz and zRy .

In the model structures for the language the self-transparent believer the access relation is serial, transitive, and weakly reflexive.

Let us call this language **LST**, for “language of the self-transparent believer”.

There are counterexamples among these model structures to show that neither $B(p \supset Bp)$ nor $B(Bp \supset p)$ is a valid sentence. So the believer’s self-transparency does not consist in having either the belief, for any and all propositions p , that p is true if, or only if, she believes that p . And it is also easy to show that the self-transparent believer never falls prey to Moore’s paradox: no such believer believes any proposition of either the form $(p \ \& \ \sim Bp)$ or $(\sim p \ \& \ Bp)$.

But the more interesting area to explore is the reasoning that is correct for the self-transparent believer when she is expressing her beliefs in first-person language. What beliefs do beliefs bring along with them, logically speaking?

6.4. Inside the self-transparent believer’s beliefs

In world w there is a designated agent, and Bp is true in w exactly if p is one of the propositions believed by that agent. So let us designate the agent’s *belief set* in that

world, the set $\{p : Bp \text{ is true in } w\}$, as $\mathbf{B}(w)$. That is in general clearly different from the set of propositions that are true in w , so we can define two distinct consequence relations. The first is the usual one, and the second depicts how beliefs will logically bring other beliefs with them. I define this relation for sentences, in the language **LST** described above:

sentence p is a *semantic consequence* of set of sentences X in **LST** (briefly $X \models p$) if and only if, for all model structures M , all worlds w in M , if all members of X are true in w then p is true in w ;

sentence p is a *doxastic consequence* of set of sentences X in **LST** (briefly, $X \Rightarrow p$) if and only if, for all model structures M , all worlds w in M , if all of X is included in $\mathbf{B}(w)$ then p is in $\mathbf{B}(w)$ as well.

We know what is the logic of \models , it is **KD4C4**. But what is the logic of \Rightarrow ? To answer this we need to look at the characteristics of belief sets. And here we will see the connection of the self-transparent believer, who avoids Moore's paradox, to the subject of rule-extended logics.

- T0. If $\mathbf{B}(w) \models p$ then p is a member of $\mathbf{B}(w)$.
- T1. If Bp is in $\mathbf{B}(w)$, then p is in $\mathbf{B}(w)$.
- T2. If p is in $\mathbf{B}(w)$, then Bp is in $\mathbf{B}(w)$.
- T3. Neither $(Bp \supset p)$ nor $(p \supset Bp)$ belongs to all belief sets for all sentences p .
- T4. Neither $(Bp \ \& \ \sim p)$ nor $(\sim Bp \ \& \ p)$ belongs to any belief set.

To put it more bluntly: in the logic of \Rightarrow , which is an extension of **KD4C4**, there are two new derivation rules

from p to infer Bp ,
from Bp to infer p ,

but no corresponding logically true conditionals. So the Deduction Theorem fails in that logic: it is non-trivial *rule extension* of **KD4C4**. In fact, that logic, the complete logic of doxastic consequence, is the following logical system:

- A1. If p is a theorem of **KD4C4** then $\vdash p$
- R1. $p, (p \supset q) \vdash q$
- R2. $p \vdash Bp$
- R3. $Bp \vdash p$

6.5. Counterexamples to Natural Deduction Rules in the logic of belief

In the logic of self-transparent belief we have therefore the direct inferences sanctioned by

$$\begin{array}{l} p \vdash Bp, \\ Bp \vdash p \end{array}$$

and we keep all the theorems and direct inferences that Frege and Russell gave us in sentential logic, as well as all those of the normal modal logic KD4C4. Familiar natural deduction rules do remain valid, *provided they are not used in combination with the new rules* that extend the classical system.

Here are some counterexamples to the validity of certain natural deduction rules in the logic of self-transparent beliefs.

Disjunction Introduction

- 1] It is raining in Peking \vdash I believe that it is raining in Peking
- 2] It is not raining in Peking \vdash I believe that it is not raining in Peking
- 3] Either it is raining in Peking or it is not raining in Peking \vdash Either I believe that it is raining in Peking or I believe that it is not raining in Peking
- 4] \vdash Either I believe that it is raining in Peking or I believe that it is not raining in Peking

If this were valid, agnosticism would be impossible. We have to disallow the move from the first two lines to the third, to block this ‘proof’.

Reductio ad Absurdum

- 1] $(Bp \ \& \ \sim p)$ *assumption*
- 2] Bp *from 1]*
- 3] p *from 2]*
- 4] $\sim p$ *from 1]*
- 5] $\sim(Bp \ \& \ \sim p)$ *from 1] – 4] Reductio*
- 6] $(Bp \supset p)$ *from 5]*

but we saw above that this is not a theorem, because $B(Bp \supset p)$ is not a theorem of KD4C4. But $(Bp \ \& \ \sim p)$ is not true on any admissible valuation.

Conditional Proof

- 1] I believe that God exists *assumption*
- 2] God exists *from 1]*
- 3] If I believe that God exists then God exists *from 1] – 2]*
- 4] In fact, I do believe that God exists *premise*
- 5] God exists *from 3], 4]*

But as we all know, there is no valid inference from belief to truth!

Postscript

In the landscape of logics beyond the deduction theorem, that I exhibited here, there are clearly pairs of logics that are the same in that they share both the theorems and the direct inference relation, but different in that certain *inference to inference rules* are valid in one but not the other. This shows that we cannot identify a logic with a direct inference relation. But can we take it that two logics are identical if they share the direct inference relation and have the same inference to inference rules admissible?

The answer is no: the landscape here depicted is itself a small region in a larger one. For this sort of inquiry has been extended very far in recent years (see especially Porter 2021).

References

- Carnap, R. 1952. Meaning Postulates. *Philosophical Studies* 3: 65–73.
- Gentzen, G. 1934-5. Untersuchungen über das logische Schliessen. *Mathematische Zeitschrift* 39: 176–210, 405–431.
- Herbrand, J. 1930. Recherches sur la théorie de la démonstration. *Travaux de la Société des Sciences et des Lettres de Varsovie, Classe III*, 3: 128pp. Also published in the series *Thèses de l'entre-deux-guerres*, 1930. http://www.numdam.org/issue/THESE_1930__110__1_0.pdf
- Hintikka, J. 1962. *Knowledge and Belief*. Ithaca, NY: Cornell.
- Jaskowski, S. 1934. On the rules of suppositions in formal logic. *Studia Logica* 1: 1–32.
- Kneale, W. & Kneale, M. 1962. *The Development of Logic*. Oxford: Clarendon Press.
- Moore, G. E. 1942a. A Reply to My Critics. In: P. A. Schilpp (ed.), *The Philosophy of G. E. Moore*, p. 533–677. Evanston, IL: Northwestern University.
- Moore, G. E. 1942b. Russell's Theory of Descriptions. In: P. A. Schilpp (ed.), *The Philosophy of Bertrand Russell*, p. 175–225. LaSalle: Open Court.
- Porter, B. 2021. Supervaluations and the Strict-Tolerant Hierarchy. *Journal of Philosophical Logic*. <https://doi.org/10.1007/s10992-021-09624-6>
- Prior, A. N. 1960. The Runabout Inference-ticket. *Analysis* 21: 38–39.
- Thomason, R. H. 2011. Some Limitations to the Psychological Orientation in Semantic Theory. *Journal of Philosophical Logic* 40: 1–14.
- Van Fraassen, B. C. 1968. Presupposition, Implication, and Self-Reference. *Journal of Philosophy* 65: 136–152.
- Van Fraassen, B. C. 1971. *Formal Semantics and Logic*. New York: Macmillan.
<http://www.princeton.edu/~fraassen/Formal%20Semantics%20and%20Logic.pdf>
- Van Fraassen, B. C. 2011. Thomason's Paradox for Belief, and Two Consequence Relations. *Journal of Philosophical Logic* 40: 15–32.

- Van Fraassen, B. C. 2021. Logic of a Self-Transparent Believer. *Filosofiska Notiser* 8(1): 11–25.
<https://filosofiskanotiser.com/Fraassen.pdf>
- Von Neumann, J. 1927. Zur Hilbertschen Beweistheorie. *Mathematische Zeitschrift* 26: 1–46.
- Whitehead, A. N. & Russell, B. 1910. *Principia Mathematica*. Volume 1. Cambridge: Cambridge University Press.

Notes

¹I say “essentially” for initially the axioms were specific sentences, and there was therefore the additional rule of Substitution. After von Neumann (1927) introduced axiom and rule schemata, so that any single axiom was replaced by infinitely many of the same syntactic form, the rule of Substitution was superfluous.

²For a three-valued logic with 1, N, 0 let 1, 0 behave in the classical manner, and for **T** as described, but assign N to any complex sentence in which one of the components has value N (‘a single jot of rat’s dung spoils the soup’, as Alan Anderson was wont to remark). But see also (van Fraassen 1968 and 1971, Ch. V, 4)

³As one of the participants in the congress pointed out, Thomason did not formalize belief with a connective but with a predicate applicable to sentences, in the style of Montague’s results about syntactic formalizations of modal discourse. My discussion here applies there *mutatis mutandis* (see van Fraassen 2011).

⁴Principle 2 has appeared in the literature, though rarely, but with a different interpretation, such as that it characterizes a ‘stable’ reasoner, in the sense that suggests a commitment not to doubt one’s own beliefs about what one believes. Taking modal statements as statements of fact, however, 2] states factually that the agent’s beliefs about what they believe are true, and that is the reading here maintained.

⁵Note: I will leave the reference to a specific truth-value assignment implicit here and below. The careful discussion in which that reference is always explicit, and all other technical details, are found in van Fraassen 2021.