Logic and Logical Philosophy
Volume 9 (2001), 199–212

Krzysztof Wójtowicz

SOME REMARKS ON HARTRY FIELD'S NOTION OF "LOGICAL CONSISTENCY"

Introduction

In this article the notion of "logical consistency" in the sense introduced by Field in [Field 1991] is discussed. Field argues, that it is possible to introduce the notion of "logical consistency" as a primitive metalogical notion, which is independent from model-theoretical and proof-theoretical notions. In this article I want to indicate some difficulties of this standpoint, and to identify some hidden presuppositions. I also would like to show, that introducing this concept (at least in the form considered by Field) does not really support Field's argumentation against realism. I confine myself to the analysis of Field's standpoint, and will not discuss the more general issue, whether it is possible to treat the notion of "logical consequence" as a primitive notion.

Field's standpoint is motivated by his discussion of Quine's indispensability argument. Field considers this argument as the only one serious and worth discussing argument in favor of realism, but rejects its premises. In Field's opinion, mathematics is really dispensable, as it plays only an auxiliary role in empirical theories. Scientific facts, which can be proved using mathematics can also be proved in nominalistic versions of these theories (cf. [Field 1980]).

In Field's argumentation, the notion of "conservativeness" plays a crucial role. Let N be a purely nominalistic ("synthetic") physical theory, S – a mathematical theory, and $N + S$ – the physical theory enriched with the mathematical tools of S. In Field's opinion, $N + S$ is conservative over N, which means, that no new consequences (formulated in the language L_N) can be found using $N + S$.

The notion of "conservativeness" can be understood in two ways:

(a) $N+S$ is syntactically conservative over N, if each sentence $\sigma \in L_N$, which is a theorem of $N + S$, is also a theorem of N. (I.e. $\mathrm{Cn}(N + S) \cap L_N = \mathrm{Cn}(N)$, where Cn is the operator of syntactic consequence).

(b) $N+S$ is semantically conservative over N, if each sentence $\sigma \in L_N$, which is the semantic consequence of $N + S$, is also a semantic consequence of N.[1]

The notion of "semantic consequence" is defined in terms of models, the notion of "syntactical consequence" it terms of proofs.

In [Field 1980], Field uses mainly the semantic conservativeness principle. To justify it he uses some metamathematical theorems, which concern classes of models for the theories in question. But this procedure enables one to object, that Field uses realistic notions in the metatheory (i.e. notions of "models", "classes of models", etc.).[2]

Field claims, that this difficulty can be eliminated, if a primitive notion of "logical consistency" is introduced. Field explicitly writes about finding "a modal surrogate of model theory and of proof theory" [Field 1991, 1], which will make it possible to justify the claims about logical consequences (i.e. facts of the form "φ is a logical consequence of T") without making use of the platonistic metatheory. In particular, it will make it possible to justify Field's claims about the conservativeness of mathematics.

The notion of "logical consistency" is — in Field's opinion — a notion, which cannot be reduced to the notions of "semantic consistency" or "syntactic consistency". It is independent from model-theoretic or proof-theoretic notions. The fact, that T is logically consistent need not be (in the general case) equivalent to the fact, that T has a model (or to the fact, that there is no proof of a contradiction within T). Nevertheless, both these facts make the notion of "logical consistency" more precise. The following facts describe the relations between these notions:

($*$) If T has a model, than T is logically consistent.

($**$) If T is logically consistent, than T proves no contradiction.

[1] Models for N and for $N + S$ usually are of different signatures, as the extralogical terms of N and $N + S$ are different, but we consider their reducts.

[2] It should be noted, that even the syntactic conservativeness principle needs the notion of "the class of all proofs", which is also a part of the "platonistic machinery".

After introducing the notation $\Diamond T$ = "T is logically consistent", these facts can be written down as:

($*$) T has a model $\Longrightarrow \Diamond T$,

($**$) $\Diamond T \Longrightarrow T$ does not prove a contradiction.

In the case of first-order logic, due to its completeness, the following equivalence is true:

$$T \text{ has a model} \iff T \text{ does not prove a contradiction.}$$

In that case, $\Diamond T$ is equivalent to the usual consistency, i.e. to the fact, that T has a model (this is a kind of "squeezing argument", which Field attributes to Kreisel). But, in Field's opinion, this fact of extensional identity is a quite accidental fact (he calls it "*accident of first-order logic*" [Field 1991, 4]), and the notion of "logical consistency" cannot — in the general case — be reduced to the notion of "having a model" or the notion of "not proving a contradiction". Nevertheless, the meaning of $\Diamond$ can be made more precise using both these notions.[3]

The discussion of the indispensability argument is not the only motivation for introducing a primitive notion of "logical consistency". In Fields opinion, it is quite natural in itself, and should be accepted even by a realist.

In this article, two groups of problems will be discussed:

(i) Some conceptual difficulties connected with the notion of "logical consistency".

(ii) The possibility of justifying the fictionalists thesis using this notion.

1. Field's arguments for the "primitiveness" of $\Diamond$

Field claims, that defining the notion of "logical consistency" in terms of models (in the Tarskian manner) is not appropriate and leads to unwanted consequences. He supports his thesis by the following arguments:

[3] Field makes use of Kreisel's argumentation, according to which we have an informal understanding of a primitive predicate Val, which means, that a certain formula is satisfied in all structures (not only in the set structures). It is formally represented by the predicate V, which says, that the formula is true in all set structures. Completeness shows, that they are coextensional, which gives more information about Val, but "that does not mean, that Val was vague before" [Kreisel 1969, 91]). Model-theoretic and proof-theoretic notions make it more precise, but its meaning is given independent of them, and Val cannot be reduced by definition to these notions.

Argument 1. Let T be the set of truths (expressed in the language of set theory) about the set-theoretic universe. If T is true, it should be consistent. But that means, that (under the standard interpretation) it should have a model M. But this model M is certainly not the entire set-theoretic universe V, but a set of the form (M, E), where $E \subseteq M^2$ is the interpretation of the predicate of set theory.[4] That means, that T is consistent not because of the fact, that it is true, but because a certain model for T exists, which need not really resemble the whole universe (i.e. $E \subseteq M^2$ need not be the "genuine $\in$"). So, in order to justify the consistency of a TRUE theory T (which describes our world), we have to rely on the existence of a model, which does not have the full set-theoretic reality in its domain (and in which the interpretation of "$\in$" does not look like membership). So the truth of T is not *prima facie* a guarantee for its consistency.

Argument 2. A similar difficulty arises in the case of logical consequence: if φ is false in the world, than S (the theory of the world — i.e. a variant of set theory) should not logically imply φ. None of these models is a "full" model, i.e. no model is identical with the whole universe, so no model needs to "mirror" the universe. So it might happen, that all models for S are models for φ, i.e. $S \vDash \varphi$ (in Tarski's sense). But then S logically implies φ, in spite of the fact, that in our world W, S is true and φ is false.

Of course, it cannot happen in the case of first-order logic (due to its completeness), but this is just an accidental feature of first-order logic, which cannot be "transferred" onto the notion of "logical consistency" for other logics. It means, that Tarski's account of "logical consequence" only works for first-order logic — and this is only "by accident".[5]

Argument 3. Let T = ZFC + "there is no inaccessible cardinal".[6] The model for T is V_λ, where λ is the least inaccessible cardinal. But in this case, inside V_λ there is no model for ZFC^2 (hence there is no model for T). So ZFC^2 is a true description of the world V_λ, but it has no model in this world, so it is not Tarski-consistent. But the explication, which allows true theories to be inconsistent is surely inappropriate.[7]

---

[4] V is not a model, as it is not a set. "From the point of view of V", V itself is not an object, and does not formally exist. The models for T are sets, which are elements of V.

[5] Let CnLog denote the "logical consequence operator" in the sense of Field. The meaning of this term is apparently given by: $\varphi \in \mathrm{CnLog}(\Phi)$ iff $\sim \Diamond(\Phi + \neg\varphi)$.

[6] ZFC^2 is second-order Zermelo-Fraenkel set theory with choice.

[7] Of course, it cannot happen in the case of first-order logic, but this — according to

2. The discussion of Field's notion of "logical consistency"

$\Diamond$ is a primitive term, but its meaning can be clarified by establishing some connections between $\Diamond$ and model- and proof-theoretic notions.

In particular, Field explains the meaning of $\Diamond$ in the following way: if S is "some reasonable finitely axiomatized set theory" ([Field 1991, 12]), then:

$$\Diamond(S^* + \exists M \vDash \varphi) \iff \Diamond\varphi\,, \tag{1}$$

or

$$\Diamond(S + \exists M \vDash \varphi) \Longrightarrow \Diamond\varphi\,. \tag{2}$$

In both cases we speak about certain connections between the logical consistency of the sentence φ, and the logical consistency of a theory, which states, that φ has a model. This explanation gives rise to several problems, which will be discussed in the sequel.

Problem 1. Field, in his thesis $\Diamond\varphi \iff \Diamond(S +$"$\exists M \vDash \varphi)$", uses the notions of "model" and "satisfaction". I will argue, that the notion $\Diamond$ is not given in a primitive way, but that sentences of the form $\Diamond\varphi$ are really abbreviations for statements about the consistency of some metatheoretical sentences. The notion of "logical consistency" is really given in the context of a certain metatheory, which serves as a background.

Every sentence φ is formulated in a certain language L. We are interested in the logical consistency of φ. We must consider the problem, whether it is necessary to use the notion of "satisfaction for L" (which is defined in the metatheory S) to justify the statement $\Diamond\varphi$, or whether it is possible to justify this fact without making use of the notion "$\vDash_L$". In particular, we have to resolve the problem, whether $\Diamond\varphi$ is given in an autonomous way and expresses certain "primitive" fact about φ, or whether $\Diamond\varphi$ is just a notational abbreviation for the sentence $\Diamond(S +$ "$\exists M \vDash \varphi$") (which is the proper explication of $\Diamond\varphi$).

In particular, we have to decide, for which of languages the operator $\Diamond$ can be applied. In other words, for which languages is the notion of "logical consistency" given as a primitive notion?

As we speak of "LOGICAL consistency", the problem, which systems should be considered as purely logical arises. Are the languages $L(Q_H)$, $L_{\omega 1\omega}$, $L(Q_0)$, etc. "purely logical"? It should be noted, that Field speaks of $\Diamond$

Field — is just an accidental feature of first-order logic and does not change the inadequacy of the model-theoretic explanation of the notion of "logical consistency".

not only in the context of first-order logic, but also in the context of second-order logic, which is a quite strong system. But in that case we should be allowed to apply the operator $\Diamond$ to weaker systems as well — in particular to the languages $L(Q_H)$, $L_{\omega 1\omega}$, $L(Q_0)$. But then several problem arise, which is illustrated by the following examples:

(A) Consider any arithmetical sentence $\varphi \in L_{\mathrm{PA}}$ (or the theory PA). If we accept the explication given above, $\Diamond(\mathrm{PA})$ means really, that the theory $S +$ "$\exists M \models \mathrm{PA}$" (i.e. set theory S with the additional postulate, that there exists a model M for PA) is logically consistent.

(B) Consider the theory $T = T_{\mathrm{lin}} + \exists y Q_0 x(x < y)$ formulated in the language $L(Q_0)$ (T_{lin} is the theory of linear order). When should T be considered logically consistent? Which criteria are in use here? If we accept the explication given above, $\Diamond T$ depends on $\Diamond(S +$ "$\exists M \models T$"), which of course depends on the metatheory S.

To accept the fact, that $\Diamond T$ depends on $\Diamond(S +$ "$\exists M \models T$") (in fact, it is either a corollary of it or is equivalent to it), we should assume, that there is a semantics for T, and that it is defined and described in the metatheory S (in particular, the notion of "model for T" is defined in S). The problem of logical consistency of T is then reduced to the problem of the logical consistency of $S +$ "$\exists M \models T$".

(C) Similar examples can be given for logics with Henkin quantifiers, for infinitary logics etc. When should the sentence $Q_H\varphi(x, y, x', y')$ (where Q_H is the Henkin quantifier) be considered logically consistent (i.e. when are we willing to accept $\Diamond Q_H\varphi(x, y, x', y')$)? Is there a primitive notion of "logical consistency for $L(Q_H)$? According to Field's explanations, $\Diamond(Q_H\varphi)$ depends on $\Diamond(S +$ "$\exists M \models Q_H\varphi$"). But this requires appealing to the metatheory S, where the notion of satisfaction for $L(Q_H)$ is defined.

(A), (B) and (C) show, that the problems of logical consistency of PA, of the sentence $Q_H\varphi$, etc. are variants of the problem of logical consistency of the metatheory S (with the additional hypothesis about the existence of appropriate model).[8]

S plays the role of a metetheory, in which model-theoretic and proof-theoretic notions for the logics in question are defined. What are the criteria for choosing a particular theory S as the "canonical metatheory"? What are the criteria for defining the semantics for the logics in question in the metatheory? We want to define the semantics in such a way, that it fulfills some methodological conditions and fits our intuitions (like the standard

[8] Of course, if S implies the existence of such models, this hypothesis is superfluous.

Tarski definition for elementary logic)[9]. This requirement should be fulfilled not only by the elementary logic, but also by every strengthening of elementary logic (like logics with additional quantifiers or infinitary logics, and also generalizations of a different type, like modal, temporal, dynamic logics, etc.). But if we accept the thesis, that the problem of logical consistency of a certain L-theory T is well-posed only if a certain "canonical" metatheory is chosen (in which $\models_L$ is defined), then the problem of the logical consistency of T is really a special case of the problem of the logical consistency of the theory $S + \exists M \models T$.

Should we accept $\Diamond T$ only if it is possible to define a model for T in an appropriate metatheory S? What is the role of S? Is our understanding of $\Diamond T$:

(A) independent of the model-theoretic facts, which are given in the metatheory S or is it

(B) just a reformulation of metatheoretical facts, which express the existence of appropriate models?

Which notions are more fundamental from the epistemological point of view?

According to (a), $\Diamond T \iff \Diamond(S + \text{“}\exists M \models T\text{”})$ in fact only identifies certain connections between the notion of "logical consistency" applied to T, and the notion of "logical consistency" applied to the metatheory S. But accepting $\Diamond T$ does not even require, that we have a knowledge of these connections, as $\Diamond$ is a primitive notion.

According to (b), $\Diamond$ can be considered primitive only at the level of S, i.e. our notion of $\Diamond T$ is derived from the notion of "logical consistency" for the metatheoretical level S.

(B) seems more plausible to me. I think, that we justify the consistency of PA or $T_{\text{lin}} + \exists x Q_0 y(y < x)$, etc. (even on the intuitive level) using the fact, that there is a structure, which serves as a model. I do not claim here, that the "feeling" of consistency of e.g. PA requires the explicit use of certain theorems of ZFC (which state the existence of ω). It is rather the case, that thinking about the consistency of PA we really think of the existence of a certain structure, in which the axioms of PA can be interpreted, and we think of PA as the description of this structure. In the case of other theories, our use of set-theoretic notions in the metatheory is even more clear. So, even if

[9] It is obvious, that just an arbitrary relation $\models$ could be defined e.g. in such a way, that every sentence is $\models$-true in every structure. In that case every theory has a model. This is not our aim.

we are in the position to have an intuitive insight into the logical consistency of theories, these intuitions can only be applied on the level of S. The statements $\Diamond$PA, $\Diamond T$, $\Diamond\varphi$ etc. are in fact only notational variants of sentences of the form $\Diamond(S + \text{“}\exists M \vDash T\text{”})$.

In this case the problem of the status of S becomes essential. The choice of a particular S is connected with the acceptance of certain set-theoretic assumptions as more plausible than others. This will have consequences for the model and proof theory of the investigated languages (e.g. the completeness of elementary logic depends on the axiom of choice in the metatheory). Of course, the existence of the model for $T = T_{\text{lin}} + \text{“}\exists y Q_0 x(x < y)\text{”}$ is not a corollary of every possible set theory S formulated in the language $L(\in)$ as it requires the existence of an infinite set and a relation with certain properties. Similarly: $\Diamond$PA is not compatible with ZFC$\setminus$Inf $+ \neg$Inf. Considering $\Diamond T$ as plausible depends on the choice of the metatheory, and $\Diamond T$, $\Diamond$PA, $\Diamond\varphi$ etc. are derived from metatheoretical statements of the form $\Diamond(S + \exists M \vDash T)$, etc. (which are epistemologically prior). This leads us to the problem of the choice and epistemological status of S.

Problem 2. From the above considerations is follows, that the sentences $\Diamond(T_{\text{lin}} + \exists y Q_0 x(x < y))$, $\Diamond(Q_H \varphi(x, y, x', y'))$ etc. really express certain metatheoretical facts. In a sense, there are many operators $\Diamond_L$ (one for each language L in question), which mirror certain “ramifications” of the notion of “logical consistency”, but one of these operators is really fundamental. Field's principle should therefore be formulated as:

$$\Diamond_S(S + \exists M \vDash \varphi) \iff \Diamond_L \varphi\,,$$

where $\Diamond_L$ is the operator of logical consistency for L ($\varphi \in L$), and $\Diamond_S$ is the operator of logical consistency for S.

I argued, that $\Diamond(T_{\text{lin}} + \exists y Q_0 x(x < y))$, $\Diamond(Q_H \varphi(x, y, x', y'))$ are really accepted, because we are able to think of certain structures, defined in the set theory S. $\Diamond T$ really means $\Diamond_S(S + \exists M \vDash T)$. Its meaning is given *via* S, and it is in fact an abbreviation for $\Diamond_S(S + \exists M \vDash T)$.

What is the status of S? Is it a “fundamental” level or should we think of $\Diamond_S$ in a similar way we think of $\Diamond_L$? Consider the following two possibilities:

(1). $\Diamond_S(S + \exists M \vDash \text{PA})$ should be explained using the meta-metatheory, i.e. by a reduction to a deeper level:

$$\Diamond_S(S + \exists M \vDash \varphi) \iff \Diamond_{S_1}(S_1 + \exists M_1 \vDash_1 S + \text{“}\exists M \vDash \varphi\text{”})\,.$$

where S_1 is the metatheory for S, $\vDash_1$ is the satisfaction relation for S, defined within the metatheory S_1, and M_1 is a model for the theory $S + \text{“}\exists M \vDash \varphi\text{”}$.

But if we accept this point of view, we are threatened by an infinitary regress. In this case the notion of "logical consistency" lacks any precise sense — we will always be compelled to claim, that it can be reduced to a deeper level. This is an idle point of view.

(2). There is a "fundamental" level for $\Diamond$. What is this level? It depends on the choice of the metatheory S, and the meaning of $\Diamond$ depends on this choice. But if the sense of the term "logical consistency" depends on the choice of S, it is quite difficult to uphold the thesis, that it is a primitive and intuitively given notion. It would be necessary to justify, that exactly one of the possible metatheories is the "genuine" one. But what are the arguments for choosing ZFC^2, or MK or BG (or any other theory, like $\mathrm{ZF}+\mathrm{AD}$) as S? It should be noted, that the possibility of formulating arguments (1) and (3) (which are intended to show the inadequacy of the model-theoretic account of logical consistency) depends on the choice of S.[10] Moreover, if $\Diamond$ applies to a particular set theory only, it is not quite clear, why it should be considered a purely logical notion.

PROBLEM 3. In his argumentation, Field uses the notion of "truth in the universe". How should this notion be understood, and on which assumption is the use of this notion based?

Field's arguments (1) and (3) have the following form:

1. Consider T, which describes the world W. Of course T is true in W.
2. The world W has the property, that INSIDE W there is no model for T.
3. So T is logically inconsistent.
4. But that means, that the model-theoretic explication of logical consistency is bad, as there is a true, but logically inconsistent theory.

Field uses the notion of "truth in the world W" (where T is formulated in the language of set theory). Using this notion, he argues, that the model-theoretic explication of the intuitive, informal notion of logical consistency is not satisfactory.

In which sense is the notion "a true sentence" understood? Consider two possibilities:

POSSIBILITY 1. The truth of the sentence $\varphi \in L(\in)$ in the world W is given in terms of the satisfaction relation $\models$ which is defined for the language $L(\in)$. But this would mean, that W is a model for $L(\in)$, since only in this case we

[10] The problem arises, whether Field's claims are not simply analytical claims, i.e. meaning postulates about the terms "$\Diamond$" and "$\models$" (given in the context of the metatheory S). But this would be inconsistent with Field's claim, that the meaning of $\Diamond$ is primitive.

can think, whether φ is satisfied in W. But this simply means, that $W \vDash \varphi$, which is impossible, as we assumed, that φ has no model. So, the truth of the sentence φ cannot be given in terms of the satisfaction relation $\vDash$. So, the second possibility must be the case:

Possibility 2. The notion of "truth of the sentence $\varphi \in L(\in)$ in the world W" should be understood in a different way — independent of the semantics '$\vDash$' for the language $L = L(\in)$. We are confronted with the following situation:

2.1. A semantics (i.e. a satisfaction relation $\vDash$) for the language L is formally defined (within a metatheory S). That means, that in our world W the class $\mathrm{Str}[L]$ and the satisfaction relation $\vDash$ are given. The models $M \in \mathrm{Str}[L]$ are elements of W (or, more generally speaking, they are somehow "connected" with W). If W is given, $\mathrm{Str}[L]$ is given in a "canonical" way. W itself is not an element of $\mathrm{Str}[L]$. The expression $W \vDash \varphi$ does not make sense — but we can speak about the truth of φ in W in a different — informal — sense.

2.2. An informal notion of "truth in W" is given. It cannot be reduced to $\vDash$.[11]

In that case, every sentence $\varphi \in L$ can be evaluated in two different ways:

2.2.a. In an informal way in the world W. Here we rest on the assumption, that the notion of "truth" does not have a formal, technical sense, but can be "grasped" on the non-formal level.

2.2.b. In formal terms (using the satisfaction relation $\vDash$) in the models $M \in \mathrm{Str}[L]$ (which are given together with the world W, e.g. as its elements).

Let $\mathrm{TR}(\varphi)$ be the abbreviation for "φ is true in the world W". The problem arises, whether the informal notion of truth TR can be "modeled" using $\vDash$. Of course, several methodological assumptions concerning $\vDash$ and the notion of "model for L" must be fulfilled, in order to satisfy the pre-formal intuitions.

Field claims, that TR cannot be modeled in a satisfactory way using $\vDash$, and that the model-theoretic account has as a consequence, that there are true but inconsistent theories. His argument has the following structure:

(i) For logics where completeness holds these notions are coextensional, but this is just an accidental fact, and the models for T can be quite "unnatural".[12]

[11] If we assume, that our universe is described e.g. within ZFC^2, and we speak about "truth in the universe", than we assume, that we can understand ("grasp") the concept of "being an element of", "arbitrary subset", "power set" etc.

[12] The problem arises, whether inner models for ZFC are natural or not. Every inner

(ii) For incomplete logics it CAN happen, that there are true sentences without a model. So $\models$ is not an adequate formal representation of TR. In this context, Field considers the theory $\mathrm{ZFC}^2 + \neg\exists\,\mathrm{Inacc}$. If this theory is true (i.e. if the universe really is V_λ, where λ is the least inaccessible cardinal), then there is no model for this theory within the universe. It is true and logically inconsistent, which is not a satisfactory state of affairs.

This argument has a conditional character: IF the world has certain properties, THEN the notion of truth has no adequate formal counterpart. It is quite unclear, whether this argument is really plausible and whether its presuppositions are acceptable. This argument presupposes some strong assumptions about the structure of the universe. Loosely speaking, these assumptions state, that the world W does not contain any objects which would make it possible to represent the notion TR in a satisfactory way. The inadequacy of the formal explication of TR is a corollary of these assumptions. But it is not *a priori* obvious, that the world HAS TO BE like that. It is possible, that the structure of the universe W makes it possible to give a formal counterpart of the notion of truth TR. What exactly does it mean? If we assume, that:

(a) the notion of TR is clear;
(b) it should be possible to give a formal counterpart in terms of '$\models$';

we are led to a reflection principle:

(REF) $\quad$ for any $\varphi \in L,\ \mathrm{TR}(\varphi) \Longrightarrow \exists M \in \mathrm{Str}[L] M \models \varphi\,.$[13]

3. $\Diamond$ and the conservativeness of mathematics

Here I shall discuss the problem, whether introducing the concept $\Diamond$ (even disregarding the problems discussed above) makes it possible to formulate a conclusive argument in favor of fictionalism (which was Fields original moti-

model contains all the ordinal number, so it differs from the universe V only "in width", not "in height" (e.g. L differs from V only in that it contains constructible sets only). '$\in$' in the inner model is inherited from the universe. An inner model is not a set "from the point of view of V", so it formally does not even exist. Of course we cannot prove the existence of such a model in ZFC, but we can do it in the metatheory. What happens, if such a model is also a model for the theory of our world? Will it be considered "natural" or not?

[13] If this reflection principle is true, that it cannot be maintained, that there is a true formula without a model. In this case Field's argumentation is "blocked".

vation). Can this notion be applied to the discussion of the indispensability argument and to the problem of ontological commitments, and does it give any conclusive arguments in this discussion?[14]

To make Field's argument for fictionalism work, we need some counterpart of the conservativeness principles. The following formulation seems the most reasonable choice:

$$(\Diamond\text{-CONS}) \qquad \Diamond N \Longrightarrow \Diamond(N+S)\,.$$

It expresses the fact, that if N is logically consistent, then it remains consistent after some mathematical assumptions S are added.[15]

Does $\Diamond$-CONS together with $(*)$ and $(**)$ make it possible to justify conservativeness of mathematics?[16] This is not the case:

1. Let us assume, that $N+S$ proves $\varphi \in L_N$. Does it mean, that N proves φ? In the general case the answer is negative. Consider the following reasoning: $N+S$ proves $\varphi \iff N+S+\neg\varphi$ is inconsistent $\Longrightarrow \sim\Diamond(N+S+\neg\varphi)$ (by $(**)$) $\Longrightarrow \sim\Diamond(N+\neg\varphi)$ (by $\Diamond$-CONS) $\Longrightarrow N+\neg\varphi$ has no model (by $(*)$).

If there is no equivalence in $(**)$ (and — according to Field — it is not necessary), then the fact, that $N+S$ proves φ implies only, that $N+\neg\varphi$ has no model. But this is not — in the general case — equivalent to the fact, that N proves φ. That means, that proving syntactic conservativeness is not possible.

2. The situation is even worse in the case of semantic conservativeness. If $N+S$ semantically implies φ, that means, that $N+S+\neg\varphi$ has no model. But this fact does not say anything $\sim\Diamond(N+S+\neg\varphi)$, as there is no equivalence in $(*)$. We cannot even start the reasoning.

[14] In the literature concerning the philosophy of mathematics the idea of eliminating the ontological commitments in favor of introducing modal notions is quite often discussed. The semantics for mathematics would be then given in terms of possibilia (possible structures, possible models, possible languages, etc.). Field's idea is similar, nevertheless his $\Diamond$ is not a modal operator. He claims, that it is rather a kind of logical modality.

[15] Conservativeness means: $\Diamond N \Longrightarrow \Diamond(N+S)$. So $\sim\Diamond(N+S) \Longrightarrow \sim\Diamond N$. In particular $\sim\Diamond(N+S+\neg\varphi) \Longrightarrow \sim\Diamond(N+\neg\varphi)$ (for $\varphi \in L_N$). So $\varphi \in \mathrm{CnLog}(N+S) \Longrightarrow \varphi \in \mathrm{CnLog}(N)$, where CnLog is the informal consequence operator, defined by $\varphi \in \mathrm{CnLog}(T)$ iff $\sim\Diamond(T+\neg\varphi)$. $\Diamond$-CONS can be formulated in the equivalent version: $\varphi \in \mathrm{CnLog}(N+S) \Longrightarrow \varphi \in \mathrm{CnLog}(N)$.

[16] Here I mean one of the standard senses (which is also assumed in [Field 1980]). If we define the conservativeness of mathematics as $\Diamond$-CONS, then the problem is reduced to the problem of justifying $\Diamond$-CONS. But if we use the term "conservativeness" in one of the standard senses, then we must investigate the problem, whether $\Diamond$-CONS allows to prove any of these standard conservativeness principles.

The principle $\Diamond$-CONS in the version given above does not make it possible to justify any of the conservativeness principles. But the idea behind the notion of "logical consistency" was to make the reasoning in favor of fictionalism work. What are the possible strategies for the fictionalist?

(A) Assume the equivalences in $(*)$ and $(**)$, but at the same time maintain, that the notions of "proof" and "model" are independent of $\Diamond$, and that these notions "fit" (in the case of a complete logic) only by accident. The argumentation of the fictionalist would then have the following scheme:

(i) A primitive notion of logical consistency is given.
(ii) Mathematical theory $N + S$ is semantically conservative over the "synthetic" theory N: it is possible to prove, that $N + S \vDash \varphi$ implies $N \vDash \varphi$.
(iii) But $\exists M \vDash S$ if and only if $\Diamond S$, as these notions are coextensional.
(iv) The fact, that N+S is semantically conservative over N makes it possible to justify the principle $\Diamond$-CONS: $\Diamond S \Longrightarrow \Diamond(N + S)$. But in this principle we only speak about $\Diamond$, not about models.
(v) So the argumentation in favor of conservativeness uses only the notion $\Diamond$, which has the same sense as the notion "having a model". But that these notions are independent, and our epistemic access to $\Diamond$ is independent from model-theoretic considerations.

Of course, this kind of argument cannot be treated serious, as here model-theoretic notions are used as well as the arbitrary assumption, that the meaning of $\Diamond$ and "having a model" are coextensional, but have different meanings. Field does not give any axioms, which would make it possible to distinguish these meanings. But in this case the statement, that the notion $\Diamond$ really has a different meaning than the notion "having a model" lacks justification.[17]

(B). Introduce a new concept of conservativeness, given by $\Diamond$-CONS. But as there is no equivalence in $(*)$ i $(**)$, this new notion of conservativeness does not say anything about the semantic or syntactic conservativeness. It only states a fact about conservativeness in the sense of CnLog (defined by $\Diamond$). But it is quite implausible, that it is precisely $\Diamond$-CONS, which adequately expresses the fact, that mathematics is just an useful tool.[18]

[17] Shapiro in [1993, 457–458] expresses doubts about the thesis, that there is a notion of logical consequence given, which fits exactly the model-theoretic notion, but is given in a primitive way.

[18] Here I do not discuss the problem, how $\Diamond$-CONS can be justified. I think, that the only possibility would be to claim, that it expresses a primitive fact about $\Diamond$. But, as it speaks about empirical and mathematical theories it is doubtful, why it should be true (as $\Diamond$ should be "purely logical").

Both (A) and (B) are not satisfactory. So, even disregarding the difficulties presented in paragraph 2, introducing the notion of "logical consistency" does not give any arguments in favor of the fictionalists standpoint.

4. Summary

1. It is not clear, at what level should the notion of "logical consistency" be applied.
2. The criteria for choosing the "fundamental level" are not given.
3. Field assumes: (i) that the notion of "truth in the universe" can be grasped; (ii) that the universe is "too small" to make a formal representation of this notion possible. But a counterargument, based on certain reflection principles can be formulated. It is far from obvious, that Field's strategy of eliminating ontology in favor of epistemological assumptions is really better.
4. Even introducing the concept of "logical consistency" does not give a convincing argument in favor of fictionalism.

In short, Field's standpoint has several weaknesses, which make in not plausible.

References

Field, H., 1980, *Science Without Numbers*, Oxford: Basil Blackwell.

Field, H., 1991, "Metalogic and Modality", *Philosophical Studies* 62, 1–22.

Kreisel, G. 1969, "Informal rigour and completeness proofs", in: J. Hinitkka (ed.) *The Philosophy of Mathematics*, Oxford University Press, 78–94.

Shapiro, S., 1987, "Principles of reflection and second-order logic", *Journal of Philosophical Logic* 16, 309–333.

Shapiro, S., 1993, "Modality and Ontology", *Mind* 102, 455–481.

KRZYSZTOF WÓJTOWICZ
Instytut Filozofii
Uniwersytet Warszawski
Krakowskie Przedmieście 3
00-047 Warszawa, Poland
wojtow@mercury.ci.uw.edu.pl