Synthese (2023) 202:100
https://doi.org/10.1007/s11229-023-04297-x

ORIGINAL RESEARCH

What is PA + con(PA) about, and where?

Jody Azzouni[1]

Received: 19 August 2022 / Accepted: 4 August 2023 / Published online: 15 September 2023

Abstract

Justin Clarke-Doane offers what purports to be a stand-alone argument, relying on Gödel's second incompleteness theorem, that if we hold that PA+Con(PA) and PA+~Con(PA) are equally true of their intended subjects, then there is no objective fact as to whether PA is consistent. It is shown that the argument is fallacious, although illuminating: The fallaciousness of the argument arises from a 20th-century shift in our understanding of interpreted languages from the view—derived from our experience of language—of sentences as intrinsically interpreted to one which sharply distinguishes syntax and semantics, and treats uninterpreted syntactic forms as endowed with interpretations by models. It is shown that this syntax/semantic distinction, because it is "unintuitive" induces fallacies such as the one that Clarke-Doane's argument exemplifies.

Keywords Clarke-Doane · First-order logic · Gödel's second incompleteness theorem · Interpretations · Models · Semantics · Syntax

The present paper shows that an argument by Clarke-Doane is fallacious. In Sect. 1 we quote the argument under study and then provide a gloss of it. Section 2 introduces the contemporary (20th -century) textbook view of formal languages: that there's a sharp distinction between uninterpreted syntax—that can nevertheless be manipulated formally in derivations—and interpretations of that syntax, which are induced via models. Section 3 returns to the argument and preliminarily observes that its conclusion doesn't follow from its premises. Section 4 shows that the apparently straightforward distinction described in Sect. 2 is hard to uphold when talking about natural language. Section 5 illustrates how, apart from natural languages, we have trouble keeping this distinction clearly in mind even when thinking about formal

✉ Jody Azzouni
jody.azzouni@tufts.edu

1 Department of Philosophy, Tufts University, Miner Hall 221b, 02155 Medford, MA, USA

languages, e.g., with respect to the Gödel completeness result about first-order logic. Section 6 provides a second pass-through of the argument with the syntax/semantic distinction in mind: the key point is that contrary to what the argument presupposes, "Con(PA)" and "~Con(PA)" aren't about (and don't say) the same things *in the respective models which render them interpreted and in which they're true*. Section 7 illustrates that the subject matters induced in sentences by models can be individuated differently: just because we treat two models as about arithmetic doesn't mean that we should treat two models derived from them by the same Gödel numbering as both about the same syntactic derivation systems. Section 8 describes how the "subject matter"—arithmetic, syntax, etc.—that a model determines (a set of sentences to be about) are individuated differently from one another. Section 9 introduces the idea of "vantage points"—from within an interpreted language, as opposed to from outside any such interpreted language, where interpreted languages and their models are compared. Section 10 describes how the two pictures of language, the older one according to which sentences are intrinsically interpreted and the new one according to which sentences are given interpretations by models, can run interference with one another. Section 11 illustrates with a little history how the syntax/semantic distinction emerged, and how difficult it was to keep it straight. Section 12 briefly describes an alternative contemporary way of understanding models and syntax. On this view, the syntax is antecedently interpreted, apart from the nonlogical terms and the quantifiers: those are parameterized, as it were (given semantic values) by models. Section 13, finally, briefly summarizes the paper.

1 The argument and an interpretation of the argument

Justin Clarke-Doane (2020, 82–83) includes, as part of his argument against "instrumental fictionalism,"[1] the following remarks (italics mine):

> … if Peano Arithmetic (PA) is consistent, then so is PA+~Con(PA), where "~Con(PA)" *codes the claim that a contradiction follows from PA*. A model of PA+~Con(PA) is a model in which there is an infinitely long "proof" of a contradiction from PA. I put "proof" in quotes, because a proof must be finite. The model is wrong about finiteness …. Or that is what we would like to say.[2] But if we hold that PA+Con(PA) and PA+~Con(PA) are equally true of their intended subjects, like, say, (pure) geometry with the Parallel Postulate and

[1] See Clarke-Doane (2020, 76–77) for a characterization of this position. It is, roughly, this: Hypothesize that there is something called "concrete reality." According to the "instrumental fictionalist," for any mathematical explanation T of some empirical phenomenon, there is a nonmathematical explanation T* that is indistinguishable from T with respect to its "implications for the concrete reality."

[2] Is *this* what we would like to say: That *the model* is wrong? Although a model is how a formal sentence is given an interpretation, that's not managed by the model "saying something" that can be evaluated for rightness or wrongness. Models just *are*; they aren't right or wrong—not as "model" is used here, for semantics. *There is* a use of "model" where models have representational roles: they're modeling something else. ("Models" are so understood in the semantic view of scientific theories.) In that case (although not here), a model *can be* right or wrong. I'm going to lean into this verbal misstep further in what follows.

> geometry with its negation[3], then there will be no objective fact as to what counts as finite and, hence, no objective fact as to what counts as a proof in PA. Consequently, there will be no objective fact as to whether PA, or any theory which interprets it, including a regimented physical theory, is consistent![4]

This argument—hereafter, *the argument*—is very rapid (that is, *the argument*—if sound and valid—is ethymematic). Despite that, its import is clear enough: it's clear what the premises of *the argument* are supposed to be, and it's equally clear what conclusion is supposed to be drawn from those premises. Almost visible, so I'll claim, is a certain family of fallacies that it relies on to work its rhetorical magic. My aim here is to unearth and discuss this important family of fallacies, both as they occur in the above quotation, and elsewhere in Clarke-Doane's book; but also as they appear—widely—in informal philosophical and mathematical discussions.

I also aim to show that this family of fallacies actually reveals a significant conceptual shift in our view of the relationship of interpreted sentences to *how* they're interpreted, a conceptual shift taking place largely over the course of the 20th century that's as dramatic as the shift in our view of space and time undergone (among cognoscenti, anyway) in light of relativity. The "fallacy," that is, that I'm claiming this argument exhibits, is one that only comes to exist, as it were, once a certain conceptual shift in our view of the syntax, and how sentences are interpreted, has occurred.

Here is my interpretation of *the argument* (in what follows I go along with Clarke-Doane's way to talking about "models" being right or wrong, but I correct it with commentary in notes):

> "What we would like to say" is that PA+~Con(PA) says that a contradiction follows from PA. And so the model in which this sentence is true is "wrong," because in that model the contradiction follows via a proof that isn't finite.[5] The alternative (the only alternative) is to take the model as not wrong. (There are two choices: The model is right or the model is wrong.) In this case PA+Con(PA) and PA+~Con(PA) are equally true of their intended subjects (just like the Parallel Postulate and geometry with the negation of the Parallel Postulate).[6] But then there is no objective fact as to whether PA is consistent, since PA+Con(PA) and PA+~Con(PA) are equally true of their intended subjects, *and according to one PA is consistent and according to the other PA isn't consistent*.

[3] See note 17.

[4] The same considerations (about Con(PA)) appear later in § 6.2 of Clarke-Doane (2020), directed this time against a version of mathematical pluralism (160): "[t]he view … that any consistent … theory has an *intended* (class) model of the sort that ZF set theory is supposed by objectivist realists to have."

[5] Or rather, since models aren't "right" or "wrong," we have to say instead that: *"PA + ~Con(PA)" is interpreted wrongly by the model in which it is interpreted as true because in that model the contradiction follows via a proof that isn't finite.*

[6] So the *model* being "right" about the sentence must be rewritten as the sentence being true *of the model*—the sentence holding in the model. As we'll see in the next section, according to the contemporary view of how sentences are interpreted, this is the only notion of interpreted sentence we're got.

2 The contemporary view of how formal languages are interpreted

If we restrict our attention to PA as a syntactic object (or as a collection of syntactic objects), we've then restricted our attention to an *uninterpreted formalism*. Thus, since a domain for the quantifiers isn't specified, the constant symbols, predicate symbols, and function symbols aren't interpreted either; that is, the sentences of PA *aren't about anything*. This is even true of the derivational rules that the connectives, ¬ and &, say, *are* given in a syntactic system. As far as uninterpreted syntax is concerned, even the classical derivational rules are just ones among many others that are possible. There are no semantic constraints on what these may be like.[7]

How is the logical and nonlogical vocabulary interpreted, and therefore, derivatively, how are the sentences in that vocabulary interpreted? The now standard answer: For the connectives, a *family* of models is given *along with* semantic conditions that link otherwise uninterpreted terminology to the models—semantic conditions that hold *across* models. The nonlogical terminology and the quantifiers are interpreted on a model-by-model basis. For this reason, models are often described as "interpretations." A derivation is a *purely syntactic process* of manipulating strings of uninterpreted symbols. It becomes something *akin to* a traditional proof when the sentences of a derivation are interpreted by a model.

3 Given the distinction between uninterpreted sentences and model-induced interpreted sentences, what is the first sentence of the argument saying?

With this in mind, let's return to part of a sentence from the quotation from Clarke-Doane, above. He writes, recall:

> … if Peano Arithmetic (PA) is consistent, then so is PA+~Con(PA), where "~Con(PA)" codes the claim that a contradiction follows from PA.

But wait! When Clarke-Doane says "PA is consistent," just the way he does in this passage, he *can't be* talking about the *interpreted* formulas of PA. He can't be *saying something about the interpreted formulas of PA* because an implication of what he means by "consistent" is "has a model,"[8] and models are understood to provide the interpretations for *otherwise uninterpreted formulas*.[9] That is, to say "PA is consistent" is to say: there is a model in which the *uninterpreted* formulas of PA are given interpretations.

[7] Consider the infamous connectives of Belnap (1962).

[8] Notice that the quotation from Clarke-Doane given in Sect. 1 focuses entirely on what the model "says" that PA+~Con(PA) holds in.

[9] Or, because of the completeness theorem, *in the first-order context but not necessarily otherwise*, this is co-extensional to: a contradiction can't be derived syntactically. See Sect. 5.

4 Applying the distinction between uninterpreted and interpreted sentences to natural language

Let us apply this apparently simple (Tarskian) apparatus (Tarski (1983a))—the distinction between uninterpreted syntax and interpretations (supplied by a model, or the world-as-a-model) in terms of which uninterpreted sentences are deemed interpreted—to natural-language sentences, and let's notice how hard it is to keep straight. (I'll diagnose why we have this difficulty in Sects. 10 and 11.) We *regularly* say things like: "'John is running and Peter is running' is consistent"; but when doing so, we're usually thinking of ourselves as talking about an *interpreted sentence* (that is, we usually have a specific sentence, about a specific John and Peter, in mind). We may also say that we can't derive a contradiction from this interpreted sentence.

Applying the above distinction: In this case we're actually talking about *syntactic* consistency ("we can't derive a contradiction")—despite the fact that we're speaking of an interpreted sentence. We're pointing out that a contradiction can't be derived from this sentence via the rules of logic, ones that we use to manipulate the syntax of sentences. This is something, though, that we can *also* point out about an *interpreted* sentence: logical rules when applied to it *syntactically* won't yield a contradiction.[10]

We may then think, because we're familiar with Gödel's completeness theorem (applied to first-order logic), or independently of that, because we're thinking of "consistency" in the sense of "can be true," that we're saying "John is running and Peter is running" is *semantically consistent*. But, as soon as we say this, we can't any longer be talking about an *interpreted* sentence, especially if, as it turns out, "John is running and Peter is running" is false. This is because to say that "John is running and Peter is running" is *semantically consistent* is to say that *there is a model* in which that sentence *as a syntactic object interpreted in that model* is true. But in *that* model (which needn't be the intended model), the sentence "John is running and Peter is running" needn't be about the same things that it's about when the sentence is interpreted by the *intended* model—the one in which we take ourselves to be talking about a particular John and a particular Peter (and that they're running). "John" may instead be interpreted (in that model) by the number 3, "Peter" by the number 5, and "Running" may designate being a prime number. The sentence, that is, *isn't about the same things* as interpreted, respectively, in the two models.[11]

It's odd, but if we say (something which sounds entirely natural) that "'John is running and Peter is running', is false, but consistent,"—"consistent," in the sense of "can be true"—we're actually running together a remark about an interpreted statement, that it's false as intended (i.e., in the intended model, in this case, the actual world), with a remark about an uninterpreted statement, that the latter is consistent, has a model.[12]

[10] This is the picture of logical proof, nearly enough, that I attribute to Frege. See Sect. 11.

[11] The point being made here is neutral with respect to issues about synonymy, and particularly, Quinean disagreements about the cogency of this idea. The point is about reference: *the same things aren't being talked about*. (This note has been added to address a concern of an anonymous referee.)

[12] What perhaps we can charitably be taken to be saying when we say this in natural language about a natural-language sentence is hard to state clearly. This is that "can be true," is meant to apply to a sentence that we "take" to have the same meaning, at least in part because we take the alternative model in ques-

5 Given this syntactic/semantic distinction, what does completeness show?

"Completeness shows," it's often said, "that syntactic consistency and semantic consistency are coextensive." That's not exactly right because the items that are purportedly coextensive, as I've illustrated in Sects. 3 and 4, needn't be the same. It takes a little finesse to state exactly what should be said about consistency in light of completeness: one can restrict one's characterization of "proof" to uninterpreted formalisms, for example. Otherwise, what completeness shows must be stated with more delicacy. In practice, of course, this particular nicety doesn't matter … except when it induces a fallacy as in the case under discussion in this paper.

Here's a key observation. Crucial to appreciating what completeness results show is distinguishing the validity of interpreted sentences from the corresponding property of syntactic derivations; they aren't the same properties even if we treat them as properties of the same kinds of objects. Derivation is, regardless, characterized purely syntactically: its characterization doesn't involve *truth*. Validity, on the other hand, is directly characterized in terms of truth, although the notion of truth, so understood, is relativized to models: that is, the uninterpreted formulas aren't themselves true or false in various models (because uninterpreted formulas aren't *about* anything); rather, they're only rendered interpreted via models, and the resulting interpreted formulas, further, are rendered true or false by those same models. Slogan: *Being interpreted and having truth-values are conjoined twins born together from the same (model-theoretic) womb*—it's not that antecedently interpreted items are *subsequently* rendered (by models) as true or false.

In contemporary studies of logic and philosophy of logic, the remarks of the last paragraph are truisms—routinely introduced to students of logic right at the beginning of their studies[13]; but two points I want to illustrate in this paper are that first, they're *contemporary* truisms. They don't become *truisms* (historically speaking) until well into the twentieth century. Second, regardless, they're never *experiential* truisms: our experience of interpreted language—when, for example, we give philosophical arguments in papers—doesn't respect these truisms. I'll illustrate the first point in Sect. 10. The rest of this paper will illustrate the second point.

6 Returning to the argument and seeing how it's fallacious

Let us return to Clarke-Doane's quotation. When he writes:

tion to be one in which "Peter," "John," and "running" refer to the "same things." (Those prone to think in terms of possible world semantics, along Kripke's lines (but not along Lewis' lines) will explain this in terms of variations across possible worlds, but given a fixed—"rigid"—interpretation of the names.) I touch on how we're to make sense of this in the mathematical context in Sect. 9. I'll claim (in Sects. 10 and 11) that in speaking this way, we're thinking about the sentence in accord with an older model of interpretation, one on which sentences carry their own interpretations across models and contexts, and don't get these interpretations "externally" by virtue of a model-assignment.

[13] Although see the discussion of the Corcoran and Shapiro (1978) review at the beginning of Sect. 10.

"~Con(PA)" codes the claim that a contradiction follows from PA,

the preceding remarks about being interpreted being induced by models require us to ask: "'~Con(PA)' codes the claim that a contradiction follows from PA," when interpreted *where*, exactly? The answer, of course, is *in the standard model of PA*. Only *there*—and in models sufficiently like the standard model—can "~Con(PA)" be described as coding the claim that a contradiction follows from PA. In any model, however, in which "~Con(PA)" is *true*, the terms in "~Con(PA)" *aren't about the same things* that they're about in models where "Con(PA)" is true,[14] and so "~Con(PA)," when true, *doesn't* "code the claim that a contradiction follows from PA": it *doesn't say* what it says in the standard model. Thus (continuing my objections to what Clarke-Doane says in what I've quoted from him above), *the model isn't wrong about finiteness*. Rather, whatever "~Con(PA)" is talking about in that nonstandard model is something it's *right* about, because it's interpreted in that model in such a way as *to be true*.[15] *The models give the sentences their interpretations*, so there's no antecedent interpretation that the sentences or the words in them (e.g., "finite") have that the "model" can be wrong about. The first option we're offered in Clarke-Doane's quotation is one that contemporary logicians and philosophers of logic can't take seriously.

According to Clarke-Doane, this forces us to the following alternative:

> But if we hold that PA + Con(PA) and PA + ~ Con(PA) are equally true of their intended subjects, like, say, (pure) geometry with the Parallel Postulate and geometry with its negation, then there will be be no objective fact as to what counts as finite and, hence, no objective fact as to what counts as a proof of PA.

But there are puzzles about why Clarke-Doane thinks what follows "then" actually follows. First, we can ask: what *is* the "intended subject" of PA + ~ Con(PA).[16] This isn't important: it can be patched up on Clarke-Doane's behalf. But, second, why is

[14] The "proofs" in models of ~ Con(PA) aren't the same things as the proofs of PA. Thus: "follows" doesn't refer to the same syntactic operation, when interpreted by these two models. Clarke-Doane acknowledges the point explicitly by putting "proof" between quotation marks and explaining, as he does in the opening quotation in Sect. 1, why he uses these; nevertheless, his so-doing doesn't render *the argument* sound.

[15] Recall notes 6 and 7 and the material they're appended to: we must rewrite talk of a *model* being right or wrong to make sense of Clarke-Doane's argument.

[16] Clarke-Doane's phrase, "the intended subject," when directed at ~ Con(PA) is more than just odd. All that he can mean here is one of *any* of the models in which ~ Con(PA) is true. This isn't the case with Con(PA), of course: *that's* got an intended model of the very syntax of PA derived from the intended arithmetical model of PA via Gödel coding. But why *is* Clarke-Doane using the phrase with respect to ~ Con(PA)? Answer: because he's leaning heavily on an analogy he perceives between Con(PA) and the parallel postulate, where—in the latter case—we can (and do) say that certain statements, incompatible with the parallel postulate, have intended subject matters. I'll say more about how we talk about the parallel postulate case shortly. Meanwhile, notice that Clarke-Doane does the same odd thing with "the negation" of the parallel postulate in the first quotation from his book, given above: he speaks of the "intended subjects" of "geometry with" the negation of the parallel postulate. There are many such subject matters, actually, no particular one of which is "intended" except in specific mathematical contexts where one or another *particular* non-Euclidean geometry is meant: this use of "intended," therefore, can't be the same as the one used with respect, say, of the standard model of PA, or (for that matter) with respect to the very proofs in PA that are the topic of intended model of (the language of) PA itself.

Clarke-Doane assuming that proofs, finite or otherwise, *of PA*, are even being talked about *at all* when PA+~Con(PA) is interpreted in a model which makes it true? And if that's the case, that proofs—finite or otherwise—of PA *aren't* being spoken of by Con(PA) when it's interpreted by a model as false, *how can* anything follow about the objectivity (or not) of the consistency of PA?

Notice the point: Take an interpreted sentence S, and now reinterpret that sentence as S*, *which is about something else entirely*. (The first interpreted sentence is about the derivational system PA; the second is about something else—not the derivational system *PA*.) How is what S* says, when true or false, relevant in any way to what S says, when true?

The logical literature uniformly takes the second incompleteness theorem as a substantive result about something very specific: *PA as a formal system*—that the consistency of *PA* can't be proven *in* PA. We know what PA is—and students have usually been shown some results in PA and practiced a bit *with PA*, before they're shown the second incompleteness theorem: PA is a particular axiomatic system, with particular inference rules—*finitary ones*. And exactly *that* is the subject matter of PA+Con(PA), *when it's interpreted in the standard model*.[17]

I'll stress what I said earlier: The uninterpreted sentences of PA, when interpreted in the standard model and when interpreted in some other model, *needn't be* about the same things at all. Their respective nonlogical vocabulary items needn't even be *extensionally* equivalent: the explicit vocabulary (e.g., the successor symbol, the addition symbol …) needn't refer to the same things. Nor, via coding these arithmetic notions into ones about syntax, as Gödel famously did, need those sentences be about the same proofs or the consistency of the same proof procedures. This is especially the case for models in which, respectively, Con(PA) and ~Con(PA) are true.

Clarke-Doane writes, as I've quoted him (the italics, here, are mine):

> *Consequently* there will be no objective fact as to whether PA, or any theory which interprets it, including a regimented physical theory, is consistent!

This only *follows* if "Peano Arithmetic," "consistent," etc., are referring to the same things in these different models—more generally, if PA+Con(PA) is about same things in both models. But they aren't. There's no way to massage the considerations Clarke-Doane has raised *in the passage I've quoted* to justify his use of "consequently." The fact that the notions of consistency, etc.—that we first-order capture (if we do) *via a standard-model interpretation of Peano Arithmetic*—aren't ones that are captured when we take that formalism and (drastically) reinterpret it in a nonstandard model, shows nothing, one way or the other, about the objectivity of those notions.[18]

I've noted that the first half of the quotation describes "what we would like to say," where that something is something we actually shouldn't like to say, because

[17] We can ask *how* this is managed; this question is set aside for the purposes of this paper, although I'll say something about the issues this question raises in Sect. 7.

[18] By a concept being "objective," Clarke-Doane (2020, 32) means that the concept doesn't amount to a plurality of concepts. He wants to say, therefore, that the concept of "straight line," in particular, isn't "objective."

it presupposes that the interpretations of the uninterpreted sentences PA+Con(PA) is the same in both models (something I'm diagnosing as implicitly presupposing that the sentence has an interpretation independently of the models that are what actually give it interpretations). Just shown is that what follows "But if" presupposes the same thing—that PA+Con(PA) is about the same things in both models (in particular, *PA*)—in order to draw its conclusion. Both alternatives rely on the same false presupposition.

7 Some caveats and observations

As I've indicated, nothing per se follows about "objective" reality—mathematical or otherwise—given our interpretation of PA in one model or another, except insofar as we can't force one or another model to be how the uninterpreted PA *must be* interpreted if all our resources are first-order and supplied only by the formalism PA itself appears in. What isn't objective, if our referential resources are restricted to the powers of a first-order formalism and supplied only by PA itself, is whether PA is *in fact* interpreted by one model or the other.

There *is*, therefore, a challenge in the neighborhood of Clarke-Doane's considerations; but it's the old (and significant) one about *how* we manage to refer to intended models—or more generally, any *specific* model, using formal language tools, or (for that matter) informal language tools. That, however, *isn't* the argument Clarke-Doane is giving in what I've quoted above.

It's this *old* challenge that motivates some logicians/philosophers to think that we can interpret formalisms, *apart from specific models*, but still treated as interpreted by families of models (e.g., 1^{st}-order logic with its set-theoretically designated family of models, 2^{nd}-order logic with its set-theoretically designated family of models, etc.) as independently interpreted apart from specific models interpreting them. I'll discuss this approach further in Sect. 12. But for now, notice that those thinking along these lines might describe 1^{st}-order Peano formalisms as "pathological" because of their nonstandard models—ones which aren't isomorphic to one another. Relatedly, many philosophers have taken referential solace in the fact that the models of 2^{nd}-order Peano arithmetic *are* isomorphic to one another. But invoking isomorphism won't avoid the change-of-reference point made in Sect. 6. After all, the point there *isn't* about what all the models Con(PA) is *true in* look like (whether they're isomorphic or not): at issue is whether the models look sufficiently alike for us to describe what's in them *as the same things*—and therefore, as the sentences talking about the same things—when those sentences are, respectively, *true or false*.

One last point. One can worry that how I've described Gödel's second incompleteness theorem—that an uninterpreted set of formulas (~Con(PA)) has a model—contradicts what that result is taken to show: That Con(PA)—as interpreted (in the standard model) as true, and thus as correctly asserting the consistency of PA—isn't provable in PA. This is a substantial result laden with what, historically, was seen as shocking (and *specific*) content: We can't prove the consistency of PA—the very axiom system that logicians prove arithmetic results in (and that's the object of study via Gödel numbering of PA)—*in* PA. How does a result that an *uninterpreted formula* has a model relate to *this*? Here's how: Con(PA) isn't provable in PA if and only if (*by Gödel's completeness theorem*) there is a model in which the uninterpreted formula Con(PA) can be interpreted, and be false (although, as stressed, what that formula is about when interpreted as false isn't what Con(PA) is about when interpreted as true in the standard model).

8 Subject matters (that models induce sentences interpreted in them to be about) are individuated differently

Let us turn to another passage from Clarke-Doane's book. In the passage to be quoted below, he's noting that most professionals don't feel the axioms of group theory target an intended model, although they do feel this way about axioms for arithmetic and analysis. Clarke-Doane (2020, 38) writes:

> In [some] cases, like group theory, the axioms do not even pretend to characterize a unique (up-to-isomorphism) intended model. There is no serious question as to whether the axiom of commutativity for groups is true, for instance. But in other cases, like analysis and arithmetic, the axioms do seem prima facie to answer to such a model.

Clarke-Doane writes, regarding the impression that there are intended models that axioms for arithmetic and analysis answer to[19]—italics mine (38–39):

> Kurt [Gödel's] Second Incompleteness Theorem implies that, if standard arithmetic, Peano Arithmetic (PA), is consistent, then so is PA conjoined with (a coding of) the claim that PA is not consistent, ~Con(PA). *So, if arithmetic were like group theory, then the question of whether PA was consistent would be like that of whether the axiom of commutativity for groups is true!* … I do not just mean that PA … might be consistent relative to one logic and inconsistent relative to a wacky alternative. I mean that there would be no objective question as to whether PA is classically consistent—that is, as to whether there is a proof of a contradiction in classical logic from the axioms of PA …. Given that there is such a question … arithmetic and set theory exhibit some objectivity.

The first sentence again confounds the syntactic consistency of Con(PA) with facts about its interpretation—in particular, its interpretation in the standard model. To repeat: In the standard model Con(PA) codes the consistency *of PA*: it doesn't in models in which Con(PA) is false. But, regardless of this, how does it follow from that: whether PA is consistent is like whether the axiom of commutativity is true of groups? The answer is that Clarke-Doane is here failing to see that arithmetic, *as a*

[19] Leave aside the fact that the mathematical subject matters, analysis and arithmetic, *aren't* codified axiomatically: the theorems, presuppositions and tools *sprawl* (and that really is the word) outside any axiomatic boundaries. See Rav (1999) for many examples. The third referee gives a nice example of this with respect to the first-order group-theory axioms: In a standard course in algebra, Lagrange's theorem is shown pretty early.

subject matter, comes apart from syntactic proof, *as a subject matter.* If they do come apart, then it doesn't follow that when we move from one model in which PA as an uninterpreted formalism is true to another in which PA as an uninterpreted formalism is true, that even if we do decide to treat both of those models *as inducing the uninterpreted formulas of PA to be about arithmetic*, that we're therefore licensed to treat both of them *as inducing the uninterpreted formulas of PA (via Gödel numbering) to be about syntactic proof via PA*. That is, even if we are "pluralistic" about arithmetic vis-à-vis possible models interpreting it, that doesn't force us to be "pluralistic" about *proofs in PA* vis-à-vis the (Gödel-numbering) induced models about "syntax." I'll develop this point further in the following paragraph.

Let's say that intended-model intuitions about arithmetic *really are* unjustified. It's reasonable (let's say instead) to treat the standard model and the nonstandard ones as all *arithmetic*. Can we do the same with our notion of syntactic proof? Certainly not, if only because of how we understand Con(PA), when interpreted in the standard model (via Gödel numbering). It's describing—somewhat idealizedly—*our methods of proof in PA*. And those involve and only can involve finite proofs.

9 Thinking about alternative geometries from a vantage point outside of those geometrical frameworks

In the passage I quoted from Clarke-Doane at the beginning of this paper, he describes a parallel between the consistency of the negation of the parallel postulate (with geometry) and the consistency of the negation of Con(PA) (with PA). Let us turn to the parallel postulate directly: doing so will show how failures to keep clearly in mind which mathematical languages we're speaking from can play out more broadly in philosophical, mathematical, and logical discussions.

There are geometries—we *all* say this on one or another occasion—in which *the parallel postulate* is false. Indeed, we describe the discovery of this as a major milestone in the evolution of mathematics.[20] On one interpretation, this milestone remark is about *syntactic* formulas. Axiomatize Euclidean geometry, and then replace the parallel postulate, or other postulates in that axiomatization, with one or more of *any* of various other postulates that together with the postulates left in place are syntactically inconsistent with original axiomatic system: the result is consistent—it has models. As I've just stressed: That's a point about syntactic formulas and about the models that syntactic formulas have.

Can we transform this claim into one that's instead about interpreted formulas? Not easily—*if we're paying attention to what we're talking about and what we're saying about it*. Suppose I draw appropriate geodesics on an orange, and I say to a student: "See? A triangle bounded by straight-line segments can have two right angles." Here I'm trying to say, among other things, that a triangle bounded by three straight-line segments—*as the student has learnt to interpret "straight-line segment" from her study of Euclidean geometry*—doesn't have to sum to exactly 180 degrees.

The student, however, responds:

[20] See, e.g., Kline (1972).

> Um … *those* aren't straight-line segments you've drawn on the orange. You can't draw *straight*-line segments on an orange. That's one of the cool things about oranges. Admittedly, if you were a two-dimensional dot on that orange traveling along one of those curved paths, *you'd think* you were traveling along a straight-line segment. Luckily, we're not two-dimensional dots traveling along curves on oranges. So … this isn't a case where "triangles" composed of straight-line segments have angles that sum to greater than 180 degrees; this is a case where there are no straight-line segments *at all*.

Is this a naïve—student-like—thing to say? Hardly. Coming out of Euclidean geometry, we can think that the technical Euclidean term "straight line" refers to something specific, although it can be generalized: "geodesic" is how the Euclidean notion "straight line" *is* generalized, i.e., brought to refer to *other sorts* of curves.[21] What about nonstandard models of arithmetic? Are those our old *counting numbers*? Those *unexpected* weirdly-structured—not well-founded—sets of things?[22] It's not obvious we have to say so.

Let's focus on the geometry case to see what's going on—because the arithmetic case is the same: We're viewing alternative *interpreted* geometric formulas from a vantage point (in a language) that's apart from—*outside*—all the specific interpreted languages that the respective axioms occur in. We're speaking from a vantage point, that is, of informal mathematical discourse. In that (meta-) language—for in that language we can talk about formal languages and models of all sorts, and we routinely do—we talk about the various models of various axioms, and we make decisions

[21] Consider the history of the function concept. See, for example, the various discussions of it, and how it evolved, in Kline (1972). Strikingly, inadequate attempts to characterize the straightness of straight lines show up among the ancient Greeks. This shows that they would not accept nonEuclidean generalizations of "straight line." This is not a point about natural language; it's a point about then-current mathematical practice. There is, for example, Plato's definition of a straight line as "that line the middle of which covers the ends." Proclus: "A line stretched to the utmost"; Heron's gloss of another definition of Proclus: "that line which, when its ends remain fixed itself remains fixed when it is, as it were, turned round in the same plane." I've taken these from Heath (1956, p. 168). Heath hypothesizes that Euclid's definition is a (failed) attempt to modify one of Plato's definitions by removing references to sight. None of these, of course, genuinely exclude the generalizations the ancient Greeks were so keen to exclude. The third referee writes: "If we grant that there are straight lines in the original Euclidean plane, then why could there not be the authentic Hyperbolic plane in which there are also straight lines? Do we want to say that the parallel axiom is built into the notion of straight line? I guess the general feeling was that would only be the case if the presumed axiom would follow from [axioms] I–IV." I think the ancient geometers wanted to capture "straight line" with a *postulate*, and yes, from the modern viewpoint, a successful characterization of that would imply the parallel axiom, given the other axioms, because it would exclude curvature *in any dimension whatsoever*—this is what the student is getting at.attempt to modify one of Plato's definitions by removing references to sight. None of these, of course, genuinely exclude the generalizations the ancient Greeks were so keen to exclude. The third referee writes: "If we grant that there are straight lines in the original Euclidean plane, then why could there not be the authentic Hyperbolic plane in which there are also straight lines? Do we want to say that the parallel axiom is built into the notion of straight line? I guess the general feeling was that would only be the case if the presumed axiom would follow from [axioms] I-IV." I think the ancient geometers wanted to capture "straight line" with a *postulate*, and yes, from the modern viewpoint, a successful characterization of that *would* imply the parallel axiom, given the other axioms, because it would exclude curvature *in any dimension whatsoever*—this is what the student is getting at.

[22] See Kaye (1991).

about whether we should call all these things that show up on saddles and on spheres, and indeed, on *all sorts* of irregularly curved surfaces, "straight lines" or not.

The result, though, is that we easily slur over our use of the statement of the parallel postulate when it occurs in one or another specific language governing a specific axiomatization, and holding of a specific subject matter (e.g., model) and when we use it—also interpreted—in our informal discussion, from (as I'll call it) "outside." We slur over, that is, the following distinction: one between the parallel postulate, where "straight line" is interpreted by the various models, and, instead, when we discuss these models and languages from outside, but continue to understand "straight line" as interpreted. When doing so (slurring between languages), we say sloppy things like: "The parallel postulate is false of *these* geometries and true of *those* geometries."[23]

I *don't* intend to claim that our talk from outside isn't legitimate, although I do think we need to analyze it carefully—something which hasn't been systematically done yet. When we're speaking from outside, about a certain class of interpreted languages and models, we can certainly (and correctly) think that certain *transmodel* (transinterpretational) claims can be made: we can notice *from outside*, for example, that the "continuum hypothesis," as we understand that phrase from outside, is true of certain set theories and not of others. In *this* case, the sets in question don't change enough across models that it illegitimates "the continuum hypothesis"—so understood—being deemed true of *those* sets but false of *these* sets. And then, additionally, we can identify "the continuum hypothesis"—as stated *in* the interpreted languages—as the same phrase with the same meaning as "the continuum hypothesis" when it's described from outside the respective interpreted languages. But this *isn't* the general case. I've suggested it's really not true of "straight lines"—in that case what we're really seeing is a generalization of the notion of "straight line" to other sorts of curves.[24] And, to stress again, it's certainly not true of Con(PA).

Indeed, it isn't true of "the continuum hypothesis." Models of set theory get pretty weird. So what's possible are set theories in which the continuum hypothesis (the syntactic object) is true or false, but where we really don't want to say that what that hypothesis is about is preserved by the shift from the family of intended models to these other models. These "sets" are such strange objects that we really don't want to say—in the sense we were wondering, *Is the continuum hypothesis true of them or not*?—that what "continuum hypothesis" refers to, *one way or the other*, is exhibited (truly or falsely) by these sets. Just this sort of thing is routine (of course) in those models of the real numbers that are countable. Zermelo makes this very point, with

[23] We switch our perspective to that of the student's dot and back again by switching between thinking of the parallel postulate as interpreted in a particular model and thinking of it as interpreted across models. To repeat: It's one thing to talk about "straight lines" using the phrase in a language that allows us to refer to all sorts of interpreted axiom systems and another to talk about "straight lines" within the specific vantage points of the specific languages of these axiom systems.The third referee stresses that this is a point that Frege "made over and over again in his protests against alternative geometries. Of course, he did not use the word 'model'." Yes, and I'll touch on how we're to understand what Frege has in mind in Sect. 11.

[24] It's similar to the generalization of Euclidean 3-space to that of "manifold." Some or all of *properties* of the intended objects are generalized. Others are just dropped.

this very example ("the Continuum Problem"), claiming that it loses its meaning for Skolem. (See Moore (1980, 124) on this).

Suppose, as certain researchers in set theories are hoping, we discover a kind of mathematical structure—and an axiomatization of it—that's conservative with respect to ZFC but has as a corollary the continuum hypothesis or its negation. It can easily be that this mathematical structure—despite being conservative with respect to the axioms of ZFC—is *so* different from the intended structure of ZFC that practitioners had in mind, that it would be unwise to say it resolves the realist question of whether sets (the mathematical objects we were intending to be talking about when using ZFC) obey the continuum hypothesis. This could be the case even if the new mathematical structure—the new set of axioms—replaced ZFC among working mathematicians.

To repeat: I'm *not* saying that it never makes sense to identify what two interpreted statements across two models are talking about. This does make sense, if only because it's what we (usually) do when, in the context of informal rigorous mathematics, we're speaking from "outside" specific axiomatizations and models. In particular, this is what we do, routinely, when we, in the context of informal rigorous mathematics, discuss different axiom systems and their models and interpret certain sentences as holding (or not holding) across models *because they're talking about what we take to be the same things*. There is, however, no *bright yellow line* about when it's reasonable to interpret certain sentences as holding (or not holding) of the same things across models and when it's not. This turns (naturally enough) on how similar those models are. There isn't, that is, anything but a messy engagement with *specific mathematical subject matters* that will tell us when we should and shouldn't identify the interpreted statements across models as talking about the same things.[25] It also turns on when and how we take certain notions to be generalizations of others and when we don't. Again: those focused on intended mathematical structures and on statements about those structures need to formulate carefully when the models that are used to interpret those statements are close enough to be deemed as inducing *the same interpretations* on the statements in question. A similar point applies to when we should regard a family of notions as belonging together: spatial curves of certain sorts, for example.

What kind of question is "When are models close enough to one another that we can treat sentences interpreted, respectively, in them, as "saying the same thing?""? Is this a factual question or a policy question?[26] This is a deep question (about *identity conditions*: are they factual when applied to models, specifically to kinds of mathematical objects?). I actually think some cases are as factual as anything we could wish for: PA, the derivational system we use, and any other logical system with infini-

[25] In particular, therefore, there's no reason to think that if we identify arithmetic structures as *arithmetic* across two models that it straightaway follows that we should identify Gödel-numbering derived syntactic structures across models of syntax resulting from those two arithmetic models—as Clarke-Doane does. Recall my discussion of this in Sect. 8.

[26] This question is due to the first referee.

tary derivation rules, are *not* the same kind of mathematical object.[27] Other cases, I'm sure, are irresolvable: just like identity conditions *in general*.

Last point: All the old (but still living) realist concerns about whether the continuum hypothesis or the axiom of choice is true of sets must be formulated carefully. Consistency results about the syntactic forms of these statements *aren't*—shouldn't be, anyway—what realists and Platonists are worried about. Their concern can be put as follows:

> We want to talk about certain specific sets. And we want to know whether they—*those specific sets*—obey the continuum hypothesis and/or the axiom of choice.

This is a *referential* concern (it's specific sets that realists are *referring to* that they're concerned with); and so syntactic consistency results obviously don't bear on their question except insofar as they bear on the question: "How does our axiomatic characterization of what we're talking about help pick out what we're talking about?" This is just the old challenge described in Sect. 7.

10 The old and new view of how semantics relates to syntax

I've interpreted Clarke-Doane's quotations, in the foregoing, as engaging in the fallacy of confounding uninterpreted formulas with interpreted ones (uninterpreted formulas accompanied by models that interpret them). This looks uncharitable on the sheer grounds that this is too obvious a distinction for Clarke-Doane to have been confused about. *I agree*—despite the textual evidence I've given. I want to now suggest that something more subtle is going on. The suggestion I'll develop in this and the next section is that there's an older perspective on interpreted sentences that's running interference here, and its interference explains why Clarke-Doane says what he says in what I've quoted. If we don't unearth the role this earlier view is (still) playing not only in Clarke-Doane's thinking but in everyone's thinking (if they're not alert to the possibility of the older perspective's potential interference in their own thinking), we'll diagnose the error in the argument he runs as only the simple one that I've described in the earlier part of this paper.

Methodological point about "fallacies": The diagnosis of a "fallacy" vis-à-vis a certain subject matter turns on the set of distinctions we allow ourselves to apply to that subject matter. This is treacherous *whenever conceptual change occurs*: we can lose touch with earlier concepts and distinctions and fail to realize that they're still playing a role in our thinking. This happens here because, as I indicated in Sect. 9, when we engage in standard informal mathematics, we're not speaking *within* syntactic formalisms (accompanied by models)—we're speaking in what *feels* just like

[27] Again, notice that the cogency of *the argument* turns on taking them to be the same. And, surely, that one derivational system is consistent (given, that is, the consistency of the background mathematics that Gödel's second incompleteness theorem is shown in) and that the other isn't *shows* they aren't the same.

ordinary language, and our experience of the interpretations of what we say then (implicitly) fits the earlier view that's been officially set aside.

What is this older view that, in my view, is running interference with the simple (post-Tarskian) distinction between uninterpreted formalisms and formalisms interpreted by models? A first pass at the older view is that it's one in which sentences of a language are *intrinsically interpreted*. Semantically-interpreted syntactic machinery is posited that explains (or partially explains) how the sentences of a language come to be interpreted, but the specific approach of fully externalizing semantics into model-theoretic structures that are what provides interpretations to syntax is absent.

What do I mean by (as in the new view) "fully externalizing semantics into model-theoretic structures that are what provides interpretations to syntax"? This: The model that supplies an interpretation does *all* the work in supplying interpretations. Domains are supplied to quantifiers relative to *each* model, the connectives are given certain interpretations relative to *all* the models—although, in general, this need not be: they could differ in their interpretations across models—and (again, specific to each model), interpretations are given to the nonlogical vocabulary. Syntax plays a role in allowing interpretations, as it were, to syntactically percolate up to larger units—e.g., open formulas and sentences. But this role is insufficient to fix the *subject matter* (that the uninterpreted formalisms can be taken to be about). When additional devices are added to the models to capture more fine-grained aspects of "meaning"—e.g., centered worlds, hyperintensional structure, etc., this doesn't change the general picture: the externalization of interpretation from what otherwise are purely syntactic objects.

Accompanying this picture of the externalization of interpretation is a similar externalization of truth—although the *title* of Tarski's seminal paper pushes truth into center stage and doesn't treat it as a corollary of the externalization of interpretation, as I'm doing here, and as Tarski himself does in the article with respect to the topic of model theory that he formalizes. Truth is definable or axiomatizable via the models that render sentences interpreted.

There can be debate about whether this externalization truly captures the intended interpretations of natural-language structures. Regardless, if we view mathematical practice from the vantage point of formal systems and their models, then the simple distinction between uninterpreted sentences in formalisms and ones interpreted by models is imposed, and any other picture of interpretation is set aside.[28]

Regardless: We're *experientially trapped* in the old view. And so, that's *not* how we ordinarily think of the sentences we speak of, for example, *our statements of mathematical theorems*. In those cases, we have a strong tendency to treat such sentences as being interpreted *independently of models*—interpretations that sentences (as it were) carry along with themselves apart from the models they're variously interpreted in. On the older view, sentences are *already* interpreted, and models aren't contexts that induce sentences to be interpreted but only ones in which sentences are accorded truth values according to the interpretations they already have. (Interpretation and truth value are *not* conjoined twins born together as in the new view.)

[28] I revisit this claim in Sect. 12.

In Sect. 9, I spoke of our identifying certain interpretations (using certain models) with each other, e.g., the continuum hypothesis as interpreted in *this* model with the continuum hypothesis as interpreted in *that* model. This *isn't* the idea of "interpreted sentence" I'm speaking of now—rather, it's one where this sentence, interpreted *independently of any model*, is true *here* or isn't true *there*.

Implicitly thinking of interpretation in this older way—as intrinsically associated with certain (otherwise syntactically-construed) sentences—impels one to say what I quoted Clarke-Doane as saying in the opening passages of this paper, *even if one officially knows better*. (If, in fact—this is the point—one *does* know better, not just officially.) What's important to realize is this older way is how we naturally—automatically—think of our sentences when they strike us as interpreted. We don't think: *What is this sentence about? Oh right, let's look at this interpretation mechanism that's operating (implicit domains of discourse, referential structures, contextually-generated referential constraints, etc.), and that gives it the interpretation it has*.[29]

A wonderful extended illustration of this phenomenon is how laborious it was for the pioneer practitioners of formal languages and logic to systematically enforce this simple distinction, between syntax and semantics, in their own thinking and understanding of formal languages and mathematical practice, as well as to forcefully apply it in polemical arguments with one another. I give in the next section a few illustrations of this.

11 A little history

Let's start with a somewhat prickly review. Corcoran and Shapiro (1978) complain, about Crossley et al. (1972, 83):

> Never is it emphasized that a sentence is true or false only under an interpretation and that it does not make sense to say that a sentence is true or false without indicating the interpretation.[30]

Why is it important that this (elementary) point is never emphasized in the textbook under review? Because—in the moment of speaking (as it were)—it's so easily forgotten by students learning the subject of formal systems (for the first time) and even by seasoned professionals, even today.[31] We naturally (unavoidably, in some sense)

[29] The semantic apparatus by which the natural-language sentences we speak and write are interpreted are, generally, invisible to us. We're almost entirely unaware, for example, how context fixes what our words refer to—how subtle cues in the environment fix interpretations. This is a reason why linguistics—as a science—is so hard, and why it emerged so late.The point: the older view is experientially supported by our almost total unawareness of the mechanisms that give the sentences we speak and write their meanings (Azzouni, 2013).

[30] The sensitivity to this point exhibited by the reviewers, on the one hand, and that one of them was John Corcoran, the editor of Tarski (1983b), on the other, is no coincidence in my view.

[31] For a carefully framed discussion of the distinction in a very-respected textbook, see Shoenfeld (1967, 2).

think of the sentences we use, e.g., the parallel postulate, as being interpreted independently of the models which give them those interpretations.

I'll illustrate this with a brief discussion of aspects of the Hilbert-Frege correspondence,[32] as well as a few observations about other early logicians with respect to this distinction. Regarding the Hilbert-Frege correspondence, it may be thought—by those who view that correspondence from the contemporary setting (and who have the syntax/semantic distinction in mind)—that their debate is partially over this distinction. *It's not*. Consider this famous exchange, with Resnik's glosses in italics:

> *Taking his own axioms to be self-evident and believing that it is impossible for a genuine axiom to be false ... Frege found consistency proofs superfluous. Thus he wrote to Hilbert*:
>
> It follows from the very truth of the axioms that they do not contradict each other. That requires no further proof.
>
> *Hilbert's reaction was dramatic:*
>
> … as long as I have thought, written and lectured about these matters, I have always declared oppositely: if arbitrarily postulated axioms do not contradict each other with their collective consequences, then they are true and the things defined by means of the axioms exist. That, for me, is the criterion of truth and existence.

Neither mathematician/philosopher has the subsequent Tarskian distinction in mind, at least not fully: Frege's use of "true" is, as it were, neither semantic nor syntactic[33]; Hilbert's use looks semantic. For both of them (for Hilbert, at least for the first few years of the 20th century, and during the time of this correspondence), the use of "contradict" looks syntactic *and* semantic. By this, in the case of Frege, I mean that he's understanding proof as mechanical, although mechanical *with respect to necessarily interpreted sentences* (that are exhibited, written, as a sequence of judgments).[34] In Hilbert's case, I think it would be overly charitable—at least at this point, although not after 1917, say—to assume that he's presupposing a completeness theorem.[35]

[32] I'm drawing my quotations from Resnik (1974). See that article for citations to the original correspondence.

[33] Frege, notoriously, regarded "true" as primitive and undefinable. See Frege (1956). See Asay (2013) for a superlative discussion of the history of this doctrine and for an ambitious attempt to philosophically update it.

[34] Resnik (1974, 387) writes: "…both [Frege and Hilbert] accepted the (then) revolutionary ideal of complete formalization which permits proofs to be checked mechanically without reference to the meaning of any of the symbols used. (This ideal, while already explicit in Frege, begins to emerge in Hilbert's letter to Frege. Of course, it became a cardinal point of the later Hilbert finitistic program)." Frege, however, as Resnik (1974) points out, Sect. 4 especially, vigorously opposed reassigning meanings to sentences in order to, in particular, construct independence proofs.

[35] Moore (1980, 117) speaking of Hilbert's views in 1900, writes of his assertion that consistency implies existence that "What was involved in this assertion was his fervent belief, at that time a philosophical

Moore (1980) while sketching the historical interplay between mathematical logic and axiomatic set theory also illustrates the various stumbles due to insufficiently distinguishing between syntax and semantics, of otherwise formidable logicians and mathematicians such as Schröder, Löweinheim, Skolem, Zermelo, and Fraenkel. Perhaps, surprisingly, these stumbles led to a focus on infinitary logics (which, among other things, was an attempt to syntactically axiomatize the otherwise semantically-characterized notion of the quantifier). Moore (1980, 96) writes:

> As the nineteenth century ended, the distinction between syntax and semantics was not uniformly observed nor even clearly understood (with the exception of Frege and to a lesser extent Hilbert).[36] This partial conflation of syntax and semantics occurred frequently within the Boolean tradition of logic, as developed by C.S. Peirce and Ernst Shröder. Consequently, the door was opened to an infinitary logic—one employing either infinitely long expressions or rules of inference with infinitely many premises.

Moore describes in detail how subsequent practitioners, Skolem, Fraenkel, Zermelo, etc., strikingly continued to stumble over the distinction.[37] Zermelo, for example, writes (quoted in Moore (1980, 120–121):

> Mathematics is *not* to be characterized by its objects (such as: space and time, forms of inner intuition, theories of numbers and measurement, and the like) but only, if one wishes to circumscribe it completely, by its peculiar process: the proof. Mathematics is a systematization of the provable and, as such, an applied logic; its task is the systematic development of 'logical systems', whereas 'pure logic' only investigates the general theory of logical systems. Now what does 'prove' mean? A 'proof' is the derivation of a new proposition from other previously given propositions, by whose truth its own is established through general logical rules or laws.

The detachment of (pure) mathematics as a topic from its presumed subject matter(s) I regard as exactly the right move to make. But Zermelo immediately after characterizes proof semantically, in terms of truth. This is very common, even today.

When is the distinction clearly seen and by whom? Well, there's Tarski, of course. But Moore (1980, 125) writes: "In his doctoral dissertation, which established the

dogma rather than a proposition capable of proof, that the consistency of the axioms for a concept implies the existence of the concept. As we would say now, the consistency of an axiom system implies the existence of a model." I'm dubious that he's—at this time—thinking of consistency purely syntactically.

[36] As I've already indicated, I demur with respect to Frege. To clearly recognize the mechanical nature of proof doesn't require clearly recognizing the (purely) syntactic nature of derivation.

[37] E.g., Löweinheim (Moore, 100) builds the notion of a domain (a semantic concept) into the syntactic expression. He also doesn't distinguish names of individuals from individuals or relation-symbols from relations. (The latter conflation—relations and relation-symbols—still occurs pretty frequently, although with no genuine ill-effects.) Relatedly, Hilbert and Fraenkel (114) conflate axiomatic set-theoretic conditions that appear, in effect, in the language with ones that appear, in effect, in the metalanguage. Recall my remarks, in Sect. 7 about slurring language within formalizations with language about (various) formalizations.

completeness theorem for first-order logic, Gödel exhibited a more profound understanding of the distinction between syntax and semantics—as well as their interrelationship—than had his predecessors."

Of course.

12 Another contemporary view of the relationship between formal languages and models?

Here is another way to think about formal languages.[38] Formal sentences, apart from models (and derivations as well) *are* meaningful by virtue of belonging to a formalism (e.g., 1st-order logic, 2nd-order logic). Their content, however, is "relative to a model"—the interpretations of formal sentences are parameterized by specific models. On this view, the 1st-order axioms of group theory—without reference to a specific model—*are* (at least partially) interpreted; they can reasonably be described as "group theory," where the models in which those axioms *hold* are, therefore, models *of* groups. All formal sentences are deemed as interpreted-subject-to-a-model-parameter.

As far as the old picture of intrinsic interpretation is concerned, this is a distinction without a difference. On one view, the models supply interpretations for all aspects of the syntax. On the other, they supply interpretations for the nonlogical predicates and the quantifiers, but not for the connectives which have them already by virtue of belonging to a formalism (1st-order or 2nd-order, etc.), and not for the quantifiers insofar as the range of models is specified by those quantifiers belonging to one or another formalism. As far as the earlier discussion of the fallacy in *the argument* is concerned, either view yields the same points.

We can still ask: Which is the right picture?[39] Once we leave the old view of interpretation behind, we face this *legislative* question, and I don't know of conclusive arguments except for considerations of *generality*: The most general picture is one in which pure uninterpreted syntax is juxtaposed with one or another family of semantic mechanisms. Pertinent is that the Tarskian approach is only one of *many* that are possible. Strikingly, substitutional approaches are possible, in which the semantics of the quantifiers occur via other language items instead of directly via items in the model. Also, there are Fregean-style approaches, where, for example, a symbol "&" semantically operates like "and" when sandwiched between certain syntactic items, and semantically operates like "or" otherwise.[40]

As mentioned, this particular conflict in what generalization we should understand contemporary formalisms in terms of—as involving a sharp syntax/semantic distinction between uninterpreted formalisms and interpreted ones, as I understand

[38] I'm including this discussion of this way of thinking because of comments of the third referee.

[39] An argument one might try against the pure syntax view of formalisms is that, as the third referee writes, italics theirs: "syntactic proof (without any specified meanings) is not a proof [because] a proof is a proof *of* something." This isn't compelling if only because the movement from the old view of interpretation (where "proofs" are composed of sentences that are intrinsically interpreted and intrinsically about a subject area) to the new one needs legislation vis-à-vis words like "proof." A consideration like "proofs are proofs of something" squarely relies on the old picture for its rhetorical power.

[40] Arguably, semantic rules like this are operating in natural languages.

the contemporary view of how formalisms are interpreted, or instead as involving a parameterized characterization of the interpretation of sentences of formalisms—can be tabled because on either view there's still a dramatic shift from our old notion of "interpreted sentence," one which induces fallacies.

13 Summary

I started with a deceptively-simple distinction, in formal and natural languages, between the syntax of those languages and the model-endowed semantics that interprets the sentences of those languages. I noted, next, a failure to respect this distinction in an argument due to Clarke-Doane. This failure, however, is due to how we experience sentences, formal or otherwise, as intrinsically interpreted. We must carefully guard against inadvertently doing this, especially in philosophical arguments. I should stress that this paper isn't evaluating whether the broader philosophical claims Clarke-Doane makes in the book this quotation is from can be sustained by setting aside his specific argument that's discussed here. This paper, instead, is focused on that particular argument because that particular argument is presented as stand-alone by Clarke-Doane, and it appears convincing precisely because it traffics in a transfer of same-interpretation of Con(PA), in the standard model (via Gödel numbering) in which it's true, to Con(PA) in a nonstandard model (via the same Gödel numbering) in which it's false. Howsoever we contemporaries understand the relationship between models and formalism (either the way I've described it in this paper, or in the way that I've mentioned in Sect. 12), this trafficking is as illegitimate as it is (nevertheless) common.

Acknowledgements My gratitude to Sean M. Carroll for drawing my attention to the argument given below in the spring of 2022. My thanks to Justin Clarke-Doane for correspondence on it in July and August of 2022. (My impression is that he remains unconvinced of what I say). My thanks to Robert Thomas for initiating my correspondence with Justin Clarke-Doane. My thanks to two anonymous referees whose comments enabled me to make this a much better paper.

Declarations

Conflict of interest This paper involves no conflict of interest on the part of the author.

References

Asay, J. (2013). *The primitivist theory of truth*. Cambridge University Press.

Azzouni, J. (2013). *Semantic perception: How the illusion of a common language arises and persists* Oxford University Press.

Belnap, N. D. (1962). Tonk, Plonk and Plink. *Analysis*, *22*(6), 130–134.

Clarke-Doane, J. (2020) *Morality & mathematics*. Oxford University Press.

Corcoran, J., & Shapiro, S. (1978). Critical study of: What is mathematical logic? *Philosophia*, *8*, 79–94.

Crossley, J. N., Ash, C. J., Brickhill, C. J., Stillwell, J. C., & William (1972). N.H. *What is mathematical logic?* Oxford University Press.

Frege, G. (1956). *The thought: A logical inquiry*. A.M. Quinton and Marcelle Quinton, trans.) *Mind new series*, *65*, 289–311.

Heath, T. L. (1956). *The Thirteen Books of Euclid's Elements* (2nd edition., 1 vol.). Dover (reprint), New York.

Kaye, R. (1991). *Models of Peano arithmetic*. Oxford University Press.

Kline, M. (1972). *Mathematical thought from ancient to modern times*. 3 volumes. Oxford University Press.

Moore, G. H. (1980). *Beyond first-order logic*. The historical interplay between mathematical logic and axiomatic set theory. *History and Philosophy of Logic*, *1*, 1–12.

Rav, Y. (1999). Why do we prove theorems? *Philosophia Mathematica*, *(3)7*, 5–41.

Resnik, M. D. (1974). *The Frege-Hilbert controversy*. *Philosophy and Phenomenological Research*, *34*, 386–403.

Shoenfeld, J. R. (1967). *Mathematical logic*. Addison-Wesley Publishing Company.

Tarski, A. (1983a). The concept of truth in formalized languages. In Tarski (1983b), 152-278.

Tarski, A. (1983b). *Logic, semantics, metamathematics*, 2nd edition. (J. Corcoran, ed., J. H. Woodger, trans.). Hackett Publishing Company, Inc.