Mario Piazza • Gabriele Pulcini
Editors

Truth, Existence and Explanation

FilMat 2016 Studies in the Philosophy of Mathematics

Springer

Contents

Chapter 4
Intensionality in Mathematics

Jaroslav Peregrin

Abstract Do mathematical expressions have intensions, or merely extensions? If we accept the standard account of intensions based on possible worlds, it would seem that the latter is the case – there is no room for nontrivial intensions in the case of non-empirical expressions. However, some vexing mathematical problems, notably Gödel's Second Incompleteness Theorem, seem to presuppose an intensional construal of some mathematical expressions. Hence, can we make room for intensions in mathematics? In this paper we argue that this is possible, provided we give up the standard approach to intensionality based on possible worlds.

4.1 Which Sentence of PA *Says* that PA Is Inconsistent?

Suppose that my friend Charles is bald. Hence the sentence

(1) Charles is bald

is true. Suppose, moreover, that Charles happens to be the tallest bald man (in the whole world, say). Hence the sentence

(2) The tallest bald man is bald

and (1) appear to "say the same thing"; namely, they ascribe baldness to Charles. Yet while (1) appears to convey something nontrivial, (2) is trivial: it says that a bald

J. Peregrin (✉)
Institute of Philosophy, Czech Academy of Sciences, Prague, Czech Republic

Faculty of Philosophy, University of Hradec Králové, Hradec Králové, Czech Republic
e-mail: peregrin@flu.cas.cz

M. Piazza, G. Pulcini (eds.), *Truth, Existence and Explanation*,
Boston Studies in the Philosophy and History of Science 334,
https://doi.org/10.1007/978-3-319-93342-9_4

man is bald, hence a truism.[1] How does this fact square with the intuition that the two sentences "say the same"?

Quite easily, of course. The two sentences say the same only given the fact that Charles is the tallest bald man, and this is an *empirical* fact which need not obtain. Hence (1) and (2) say the same only on the assumption that

(3) Charles is the tallest bald man

It follows that the sense of "saying the same" just considered cannot be identical with "meaning the same", i.e. with synonymy. The point is that the meaning of (1) and (2), and hence the identity or non-identity of their meanings, cannot depend on empirical facts of the kind spelled out by (3). And if (1) and (2) were to mean the same thing, then in a case where (3) were false they could not cease to mean the same thing (though there would be a sense in which they would no longer say the same).

Now consider a *prima facie* similar case. Suppose that some formal theory, say Peano arithmetic (PA), is consistent, hence the sentence

(1*) PA is consistent

is true. It follows that PA is the biggest consistent theory included in PA, hence that

(2*) The biggest consistent theory included in PA is consistent

Just as in the previous case, the two sentences appear, on the one hand, to "say the same"; namely, they ascribe consistency to PA while not being, on the other hand, really synonymous. But why are they not really synonymous? We might be tempted to give an answer wholly analogous to the previous case, namely that (1*) and (2*) "saying the same" is conditioned by

(3*) PA is the biggest consistent theory included in PA,

which is an empirical fact that need not obtain. But this would not work here: (3*) is *not* an empirical fact and it can*not* fail to obtain. If PA is consistent, there cannot be a situation in which it would fail to be identical with the biggest consistent theory included in PA.

Is it then the case that (1*) and (2*), unlike (1) and (2), *do* have the same meaning?

We may tend to say that this perhaps is not a question worth dwelling on. In a sense, (1*) and (2*) do mean the same thing, while perhaps in another sense they do not. But this cavalier approach would lead us into some difficulties, not the smallest of them being that the answer to the synonymy question is presupposed by one of the most significant findings in logic of the twentieth century, Gödel's Second Incompleteness Theorem.

As pointed out by Auerbach (1985), while the First Incompleteness Theorem is a plain statement regarding the structure of certain formal systems (if the axioms of

[1] Well, perhaps it says, over and above this, that there is one and only one tallest bald man; however, this again does not seem to be too nontrivial.

such a system include a modicum of arithmetic, then there is bound to be a sentence which is independent of them), the nature of the Second Theorem is different – it presupposes that certain arithmetical sentences have certain meanings. (Therefore, Feferman 1960 takes it to be an *intensional* problem.[2]) The Second Incompleteness Theorem says that no system of the above kind proves a theorem that says of the system that it is consistent. Hence, the very formulation of the Theorem assumes that we know what it takes for a sentence of the system to say (mean?) that the system is consistent.

At first sight, it might seem that this is not a problem and that we do not need to enter the muddy waters of the considerations displayed above. To say that a system is consistent is to say that it does not prove the absurdity (often, the sentence "$0=1$" is used as its embodiment), or that there is something which it does not prove. Hence everything we need is the provability predicate, which, in turn, is based on the binary predicate Prf that relates two numbers iff the first is the Gödel number of a proof of a theorem whose Gödel number is the second argument.

Thus it might seem that a sentence's saying that the system is consistent can be reduced to it being of the form $\neg\exists x\mathrm{Prf}(x, \ulcorner 0=1 \urcorner)$, where $\mathrm{Prf}(x, y)$ says that x is (the Gödel number of) a proof of (the Gödel number of) y. And $\mathrm{Prf}(x, y)$'s saying this in turn reduces to its relating all and only numbers of proofs with numbers of theorems they are proofs of. But this unfortunately is not the case, and we are led directly into the muddy waters that we hoped to avoid.

Consider the binary predicate $\mathrm{Prf_R}$ defined in the following way (it was, in effect, proposed by Rosser 1936):

> $\mathrm{Prf_R}(x, y) \equiv_{Def.}$ x is a proof of y and x is not greater than the number of any proof of any negation of y.

Given PA is consistent, the existence of a proof of y excludes the existence of a proof of its negation, and hence the number of a proof of y can never be greater than the number of a proof of its negation; hence $\mathrm{Prf_R}(x, y)$ iff $\mathrm{Prf}(x, y)$. It follows that in so far as saying that x is (the Gödel number of) a proof of (the Gödel number of) y amounts to nothing more than relating all and only numbers of proofs with numbers of theorems they are proofs of, $\mathrm{Prf_R}(x, y)$ says this just as $\mathrm{Prf}(x, y)$ does. But the trouble is that if we substitute $\mathrm{Prf_R}$ for Prf in the consistency sentence, we get a sentence that is *provable*.[3] Hence, either Gödel's Second Incompleteness Theorem does not hold, or $\mathrm{Prf_R}(x, y)$ does not say that x is a proof of y after all. And, as the

---

[2] A thorough discussion of the ways in which the Second Incompleteness Theorem is intensional is given by Halbach and Visser (2014a,b).

[3] This is quite obvious if, instead of $\mathrm{Prf_R}$, we use still another predicate, namely $\mathrm{Prf_C}(x, y) \equiv_{Def.}$ $\mathrm{Prf}(x, y) \wedge \neg\mathrm{Prf}(x, \ulcorner 0=1 \urcorner)$. Given $\mathrm{Prf}(x, \ulcorner 0=1 \urcorner)$ for no x, $\mathrm{Prf_C}(x, y)$ holds for the same x's and y's as Prf (x, y); while $\neg\exists x\mathrm{Prf_C}(x, \ulcorner 0=1 \urcorner)$ is $\neg\exists x(\mathrm{Prf}(x, \ulcorner 0=1 \urcorner) \wedge \neg\mathrm{Prf}(x, \ulcorner 0=1 \urcorner))$, which is obviously provable. In fact, if we consider a further modified version of $\mathrm{Prf_C}(x, y)$, $\mathrm{Prf_{C'}}(x, y) \equiv_{Def.}$ $\mathrm{Prf}(x, y) \wedge \forall x\, \neg\mathrm{Prf}(x, \ulcorner 0=1 \urcorner)$, we may return to our example (1*) vs. (2*), for we may think of (1*) as roughly capturing the sense of the standard consistency sentence based on the predicate Prf, and of (2*) as capturing that of its variant based on $\mathrm{Prf_{C'}}$ – for $\forall x\, \neg\mathrm{Prf}(x, \ulcorner 0=1 \urcorner)$ clearly amounts to the consistency of PA. (Cf. Auerbach 1992).

Second Theorem would be taken as a well-established fact, we seem to be back to pondering what does it take for an expression of a formal system to say something.[4]

4.2 Possible Worlds

Let us return to sentences (1) and (2). The fact that there is a sense in which they do "say the same" and a sense in which they do not is clear, and by now it has a standard explanation: they do "say the same" on the level of reference, while they do not "say the same" on the level of meaning. The fact that semantics must consist of two levels, such as meaning and reference, now an established commonplace, was most clearly articulated by Frege (1892a); and since his seminal analyses, almost no theory of semantics has avoided a stratification of this kind.

Why do we need the two levels? An answer follows from a consideration of what it is that language is good for. On the one hand, we certainly need language for talking about objects that are around us: the chair on which I am sitting, my computer, my friend Charles, the current president of the USA, etc. On the other hand, we need language to be very general and universal – its semantics should not depend on there being my computer or my friend Charles. Doing justice to both these requirements works in such a way that linguistic expressions have general *meanings*, which conspire together with the empirical facts to produce *referents*. Thus, the expression *the current president of the USA* has a general meaning, independent of the actual possessor of this office, but given the current state of the USA, it is able to refer to a concrete person.

Frege famously called the two levels of semantics the level of *Sinn* (sense) and that of *Bedeutung* (meaning) – a not very happy choice, for it is his *Sinn*, rather than his *Bedeutung*, that corresponds to meaning in the intuitive sense of the word. From this viewpoint, it was helpful when Carnap (1947) proposed altering the terminology and rechristened Frege's *Sinne* as *intensions* and his *Bedeutungen* as *extensions*. This is the terminology that has become standard and which lets us say that sentences (1) and (2), and especially their subjects, *Charles* and *the tallest bald man*, share the same extension while they differ in intension.

How best to get an explicitly semantic grip on the difference between intension and extension was foreshadowed already by Carnap, but came to full fruition only after Kripke (1963a,b, 1965) introduced his semantics for modal logic based on the notion of possible world. The basic idea of so-called intensional semantics[5] that grew from the Carnapian considerations regarding the nature of intensions and the

[4]We may think about a "non-intensional" content of the Second Incompleteness Theorem: for example we may take it as saying that we cannot prove $\mathrm{Pr}(\ulcorner 0 = 1 \urcorner)$ for any predicate Pr fulfilling Löb's derivability conditions (Löb 1955; cf. also the thorough discussion of Boolos 1995). But then the question is why this result should be very interesting; unless we show that there is a reason to think that fulfilling Löb's conditions amounts to being the provability predicate, it is a far cry from what is usually taken as the Second Incompleteness Theorem (see Detlefsen 1986).

[5]Put forward by Montague (1974) and others.

apparatus of possible worlds was that the intension may be seen as something like an extension relativized to possible worlds; more technically, it can be considered as a function mapping possible worlds on corresponding extensions.[6] Hence, while the extension of *the tallest bald man* is (presumably) a concrete person (my friend Charles), its intension is a function that maps any possible world on its tallest bald man (if any).

Let me call the result of this development the *possible-worlds account for intensionality*, or PWAI, for short. To many philosophers it appeared to be the ultimate solution to the problem of intensionality. However, it is usually taken to have no impact on mathematics. The facts of mathematics, it is usually assumed, are independent of the way the world is, hence they are constant across possible worlds and if we therefore assign intensions to mathematical sentences they will be constant functions mapping every possible world on one of the truth values.

Hence, given PWAI, there is no space for intensionality in mathematics. According to PWAI, all true mathematical sentences have the same intension and so do all the false ones. (This was also behind the most spectacular failures of the accounts for the so called propositional attitude reports provided by PWAI – see Bigelow 1978 or Partee 1982.) Some semanticists proposed that semantics must go "hyperintensional" (see, e.g., Lewis 1972 or Cresswell 1975).

Most working mathematicians do not seem to care – after all, when doing mathematics and using its language there is no need to poke into the details of its semantics. It is clear that for many mathematical expressions, such as e.g. "1+3" and "2+2", though they have the same extension (in this case the number 4) and from the viewpoint of PWAI also the same intension (namely the function mapping every possible world on the number 4), they do not have, intuitively, the same meaning; but it is not clear that mathematicians (in contrast to philosophers of language) need to rack their brains over this apparent clash with everyday intuition. However, we have seen that some problems, such as the Second Incompleteness Theorems, which are often understood as eminently mathematical, do presuppose a clarification of meaning far beyond simple intuition.

4.3 Consequence and Meaning

From the viewpoint of PWAI, mathematical discourse is anomalous in that it lacks the extension/intension distinction, which appears to be vital for non-mathematical discourse. Let us consider the ways in which this anomaly influences the concept of consequence and then the very concept of meaning.

What is consequence? We can say that

(4) Fido is an animal

[6] This is an oversimplification; technically, the systems of intensional logic tend to be somewhat more complicated. See Peregrin (2006).

is a consequence of

(5) Fido is a dog

and similarly we can say that

(4*) L has the smallest element

is a consequence of

(5*) L is a lattice.

What, in general, does it take for a sentence to be a consequence of other sentences?

Intuitively, A is a consequence of $A_1, \ldots, A_n$ iff the truth of $A_1, \ldots, A_n$ "forces" the truth of A, i.e. if the latter is true whenever the former are true. The "whenever" just employed indicates that this definition contains a tacit quantification: A is true in all cases when $A_1, \ldots, A_n$ are – but what are the cases over which we quantify? For (4) and (5), the cases would seem to be the possible states of the world (which may be captured by Kripkean possible worlds): all states in which Fido is a dog are bound to be states in which Fido is an animal (while in other states this need not be the case). However, this does not work for (4*) and (5*), for if L is a lattice, it remains a lattice whichever state of the world obtains.

Let me stress that just like the case of *Fido* in (4) and (5), L in (4*) and (5*) is supposed to be the name of a specific object (not just a kind of variable), for (4*) and (5*), like (4) and (5), are supposed to be sentences, not just schemata. But could we not use the Tarskian definition of consequence, which, in effect, reduces the consequence among sentences to consequence among the corresponding logical forms (where the latter obtains iff everything that is the model of the premises is also a model of the conclusion)?

Not really. The fact that (5*) follows from (4*), obviously, is not a matter of their logical forms, it is a matter of the fact that lattices are bound to have smallest elements. But could we not say that (4*)'s being a consequence of (5*) amounts to the quantificational statement

(6*) for every x, if x is a lattice, then x has the smallest element?

This is also problematic. For consider the analogue for the case of (4) and (5).

(6) for every x, if x is a dog, then x is an animal.

It is true; but it would not be generally held that its truth would guarantee the consequence. Consider the sentence

(7) for every x, if x is a featherless biped, then x is a human.

This is also true, but it certainly does not follow that *Charles is a human* is a consequence of *Charles is a featherless biped*. The reason is that the truth of (7) is *contingent*; in terms of PWAI we can say that though it is true in the actual world, it is not true in *every* possible world.

Hence we can see that a conditional of this kind guarantees consequence again only if it is true not only about the actual world, but across possible worlds. Does this hold of (6*)? Yes, but somewhat vacuously, for a mathematical sentence is true in every possible world as soon as it is true in the actual one. It follows that consequence in mathematics, if considered in this standard way, boils down to material implication; hence it also suffers from all the well-known maladies of this kind of implication.[7]

Maybe this is something we should simply accept. But what about meaning? Similar arguments indicate that furnishing an expression with an extension is insufficient for making it genuinely meaningful.

Suppose we stipulate that a word refers to the set of current humans. What have we thus made the word *mean*? *Human*? Or *featherless biped*? These are definitely two different meanings, hence our new term has to mean one or the other (or still something else). The answer is that in this way we do not make it mean anything, for conferring extension is notoriously short of conferring meaning. To make a word meaningful, we need to furnish it with an intension. Now how is it with meanings of mathematical expressions?

Mathematical expressions, of course, *may* be thought of as having intensions, but then the only intensions available are those that are constant over possible worlds and hence are uniquely determined by extensions (and hence are somewhat superfluous). Thus to furnish a mathematical expression with an extension *is* to furnish it with an intension, albeit trivially. However, this would mean that, e.g., the expressions *PA* and *the biggest consistent part of PA* mean the same thing, which does not seem to be viable.

One possible approach to this situation is to revise the concept of possible world on which we base the PWAI. A proposal, going back to Hintikka (1975), is to consider not only worlds that are "ontologically" possible, but also worlds that are "epistemologically" possible (though they are ontologically impossible). This means that, for example, in a situation where there is no proof of the consistency of PA, there is an (impossible) possible world in which PA is not consistent.

It seems to me that this proposal is on the right track – as far as the semantics of possible-worlds goes. An obvious objection, however, is that there is no way of determining what is epistemologically possible. Is it what any speaker of the language in question can hold for true? But would the ignorance and/or folly of some speakers stop short of anything – would there be anything that would be at all impossible?

Perhaps we can resort to a whole society as a measure of epistemic possibility. Perhaps something is epistemically possible iff the contrary has not been "ascertained" (proved, in the case of mathematics) by the community. According to this proposal, our community no longer allows for an (epistemologically) possible world in which Fermat's Last Theorem would not hold, but it does allow for one in which the Goldbach conjecture does not hold. But this would mean that there is also no world in which PA is inconsistent (at least if we do not want to challenge the existing proofs of its consistency) and *PA* would come out again as co-intensional with *the biggest consistent part of PA*.

[7] See Peregrin (2014, Chapter 7).

4.4 Back to Frege

Frege's approach to intensionality differs from that of his continuator Carnap in at least two respects. Frege's *Sinn*, "intension", is a *way of givenness* (*Art des Gegebensein*; see Frege 1892a) (and this is pretty much everything Frege tells us about it). And, intuitively, it would seem that specifying PA by way of *PA* and specifying it by way of *the biggest consistent part of PA* are two *different ways.*

But, of course, we must not assume that two nominally different phrases always mean different *ways of givenness* – not all different expressions need to mean different things (independently of which particular theory of meaning we subscribe to). And possible worlds appear to offer us precisely a criterion of determining when two individual descriptions are semantically different and when not: they are different if there is a possible world in which they do not describe the same entity. Is it possible that *the biggest consistent part of PA* does not describe PA? Not according to PWAI.

Frege's opinion was arguably different. For what he gives as one of his pivot examples of different ways of giving the same object is the following:

> Let *a, b, c* be the lines connecting the vertices of a triangle with the midpoints of the opposite sides. The point of intersection of *a* and *b* is then the same as the point of intersection of *b* and *c*. So we have different designations for the same point, and these names ('point of intersection of *a* and *b*,' 'point of intersection of *b* and *c*') likewise indicate the mode of presentation; and hence the statement contains actual knowledge. (Frege 1892b, p. 26)

As it is clear that the point of intersection remains the same however we vary the world, the two descriptions have the same intension (on the standard construal of possible worlds) – from this viewpoint they therefore do *not* appear as different descriptions. Hence Frege thought that two descriptions can be semantically different even if they are (according to PWAI) co-intensional.

Let us consider the idea underlying PWAI. To have, aside an extension, also an intension is to have an extension not only now, but also in various counterfactual situations – to be, as it were, "counterfactually robust". Perhaps we need not be sure what the extension of the term might be in some very exotic circumstances (what would be the extension of *human* in a world with many more or less humanoid species?), but to understand the word is to know at least how its extension moves when the circumstances shift from the current ones into their neighborhood.

Of course, if we consider changes of circumstances, we must concentrate on the way the changes affect the object or objects that constitute the extension in question – to know, that is, the rules that regulate the movements. Unless the object is affected, the change is not really relevant. Hence considering counterfactual circumstances is, in effect, considering variants of the object or objects in question and considering which of the variants still count as "the same" object or "the same" kind of objects. And this depends on what we are, and what we are not, willing to consider "the same".

PWAI is based on the notion of "ontological possibility", which is taken to be an objective matter, wholly independent of us. (In this way, the standard treatment is usually not only supposed to *account for* meaning but also to *explain it*, to reduce it to something non-semantic.) The above considerations indicate that this is doubtful. Why is it that the sentence *Fido is a dog and is not a dog* cannot be true? Because there is no possible world in which it is true? But how do we know, have we checked all the possible worlds, one by one? We know it because we know (explicitly or implicitly) the rules of our language (in this particular case the rules governing *not*), and we know the truth is excluded by these rules. Thus we do not know that the sentence is necessarily false because it is false in every possible world; rather, we know that it is necessarily false, because this follows from the rules of our language that we mastered on learning it, and we therefore know that there cannot be a world in which the sentence is true. Hence, possible worlds are a certain (useful) means of envisaging certain rules of our language, not something that would explain why the rules are such as they are.[8]

Take (1). It would be taken as a prototype of a contingent, empirical sentence – hence a sentence that is *not* true in every possible world. But why is this? Clearly it depends on what kind of entity we consider the individual Charles to be. When I point at Charles, I point at an intricately complex configuration of molecules making up an entity with a lot of physical features, including a bald head – and precisely this particular configuration cannot lack the bald head, for it would be a *different* entity if it did. The fact that we take (1) to be contingent is because we individuate individuals more loosely than the individual configurations of molecules – many more or less different configurations of this kind may count as the same individual. (Though what is and what is not the same individual is not always clear; and it is much more unclear in cases of things other than animals – *viz.* the celebrated case of the ship of Theseus.) Hence, saying that what is and is not contingent is clearly determined by "ontology" is a huge exaggeration.

The situation is not very different in the case of extension for general terms, such as, e.g., *human*. How much can the extension of the class of current humans vary while still remaining the class of humans? In contrast to the extension of an individual term, such as Charles, here we must take into account not only the variations of the extension itself, but also those of the world around it. It is clear that the extension may gain a lot of kinds of individuals in many ways differing from the current ones; and it can lose any of the current ones (if they are no longer part of the world). It is also clear that gaining individuals differing in more substantial ways (e.g. not having brains) might compromise the status of the set as the set of humans; as well as losing some current individuals (if they still are part of the world). Thus, though the boundaries of the species *human* are far from clear-cut, there are many features that obviously do not move an individual outside the boundaries and there are many such that do. And to understand the term *human* is to understand these.

Note also that we need not be able to see all the conditions that shape the boundaries of the extensions. Imagine a kind of unobvious bodily feature X without which one cannot have a brain. Thus, insofar as we would hardly take a brain-less

[8] I discussed this in great detail in Peregrin (1995).

organism to be human, the lack of X would prevent an individual from being a member of the corresponding extension, though we need not know about this lack. In general, this is a matter of the fact that we need not see all the consequences of the rules of our language. It follows that we may consider a class of organisms containing some X-less ones as admissible classes of humans.

In fact, the claim that humans have X would belong to what Kripke (1972) calls the "necessary a posteriori". (While Kant famously urged that there are not only analytic a priori and synthetic a posteriori judgments, but also synthetic a priori ones, Kripke argues that we can have, the other way around, analytic a posteriori judgments.) The criteria constitutive of a meaning of a term may have consequences we are not able to see, and hence there are properties which anything falling under the term must have in force of the meaning the term (and hence the claim that it has the property is analytic); though to find out about them might be a discovery (thus the claim is a posteriori).

The important point is that this holds both for empirical and mathematical terms. Our normal usage of the name "PA" is such that their bearer gets individuated in terms of the axioms constitutive of it. We certainly intend to refer to the unique system determined by the axioms with all its properties, but in the case of us not knowing all the properties, or in the case of not all the properties being obvious, the meaning of "PA", as embodied in its use, does not exclude some variants of the intended reference. (Unlike reference, the meaning is something we know when we understand an expression. Thus, what is contingent – in the sense of Kripkean a posteriori – cannot be part of the meaning.) Note that if this were not the case, the *meaning* of "PA" would be unknown also to many of those who would know the axioms of the system perfectly well, which would seem to be strange.

Perhaps, then, we can see (1*) as contingent, for it is a posteriori in the Kripkean sense. If PA were slightly different from what it is, it would not need to be consistent; and the slight difference would not make us say it is no longer PA (because we need not be able to see all its consequences). Then we could say that just as the fact that *Charles is bald* is contingent because we can imagine something non-bald and close to the current configuration of molecules making up Charles (enough so that we are content to call it "the same individual"), the fact that *PA is consistent* is contingent because we can imagine something inconsistent and close to PA in a strict enough sense that we are content with calling it "the same system".

This opens up a novel view of intensionality. Entities of the actual world may have "neighborhoods" consisting of entities more or less similar to them that are considered to be "identical with them". The configuration of molecules called at this moment *Charles* continues to be the same Charles even if it undergoes various kinds of changes. The set of humans of the actual world, which is the extension of *human*, remains a set of humans even after some of its members are removed and new members – of the appropriate kind – are added. And PA remains the same system providing we tamper with it only in ways which we see as not threatening its identity.

4.5 Intension as "Robustness"

What is important is that to know the intension of a term is to know in which directions its extension can be moved without ceasing to be the extension, and in which directions this is not possible. Consider the conglomerate of molecules constituting the current Charles. There are many changes that can happen in this entity that do not cause it to no longer be Charles. Indeed, almost any minor change is insubstantial; only an entity substantially different from the current one may not count as Charles. (Though even an entity very radically different from the current Charles may still count as Charles if it results from certain transformations of the current Charles – the criteria of what still counts as Charles and what no longer does are quite complicated and not always unequivocal.)

Thus, we may exploit the wisdom behind PWAI in a slightly different way than PWAI itself does: we can say that an expression is furnished not only with extension, but also with intension if it is determined to what extent it is "robust" – how various kinds of changes of the *status quo* tamper with its extension, especially which kinds of variants of the current extension are prone to take over its role. And, in contrast to PWAI, this alternative view of intensionality may also be usable in the realm of mathematics.

Consider the binary relation that links $\ulcorner x \urcorner$ with $\ulcorner y \urcorner$ iff *x is a proof of y*. It is clear that if a (number of a) proof plus a (number of a) sentence are in this relation w.r.t. a given axiomatic system, this cannot be changed by adding more axioms to the system (nor, for that matter, to another one). The only thing that can result from the extension of the list of axioms is that more numbers will be in this relation. Hence if we consider the relation links $\ulcorner x \urcorner$ with $\ulcorner y \urcorner$ iff *x is a proof of y and PA is consistent*, we can see that it behaves differently – adding axioms to PA so that it becomes inconsistent exempts any sentences from this relation. (The same holds for the more complicated version *x is a proof of y and there is no proof z of a negation of y such that z< x*).

Now given an axiomatic system, the claim *there is something that is not derivable from the empty set of premises*, if false, can be made true by subtracting axioms, but not by adding more of them to the system (and certainly not to PA, if PA is not the system under consideration); whereas if we used *x is a proof of y and PA is consistent* instead of *x is a proof of y*, this would not be the case. Hence if we defined *is consistent* as *there is something that is not derivable from an empty set of premises*, it is ok for the former derivability predicate, but not for the latter one.

If we thus accept that a necessary condition for a term having an intension is its "determinateness" in the vicinity of its actual extension (in the sense that it is clear which variants of the current extension still are its potential extensions, and which not), we have a criterion for deciding which predicates are reasonable candidates for expressing the property of consistency (and similarly for other mathematical terms). The criterion is not sharp and its applicability is limited; however, it brings forth a possible way of explicating the intuition that mathematical terms also have intensions.

Now let us return to the statement

(1*) PA is consistent

Does it hold necessarily? Consider the statement (where X is still a feature an organism must have to have a brain and such that a normal speaker is not likely to know about it):

(1**) Charles has X.

Does *it* hold necessarily? Yes, it does, in the sense that without X Charles would not be what we are normally willing to call a human being, and hence *a fortiori* would not be Charles. However, this is a piece of empirical knowledge (based on the assumption that the biological theory conditioning having a brain, and hence being human, by having X holds) – it is not a matter of the rules of our language. To fail to know that Charles has X (which, we assumed, a normal speaker does) is not to fail to understand the term *human* or the name *Charles*, it is to fail to know something about humans and especially about Charles.

Similarly (1*) does hold necessarily, for we know that the lack of consistency would be possible only when PA were a different theory. Unlike the case of X and Charles, this is not empirical knowledge, but just like in this case it is conditioned by the fact that there are no errors in the proofs of the consistency of PA. (What if all the proofs of the consistency of PA presented so far are in fact erroneous, it is only that we have missed the errors? This is, needless to say, very improbable, but certainly neither impossible in some metaphysical way, nor excluded by the rules of our language.) Hence we can say though (1*), in contrast to (1**), does not amount to empirical knowledge, it does amount, in accordance with (1**), to knowledge that is a posteriori in the Kripkean sense.

In contrast to this, there is no such possibility for (2*). However we vary PA, it not only cannot be the case that its biggest consistent part would not be consistent, but also it cannot be the case that we would have to discover that it is consistent. This, I think, accounts for both the intuitive distinction between (1*) and (2*) and the intuitive parallel between (1) and (2), on the one hand, and (1*) and (2*), on the other – both pairs get equated only via an additional nontrivial fact that is a posteriori, though in the first case in the ordinary synthetic a posteriori sense (the fact is simply empirical), while in the second case in the sense of the Kripkean necessary a posteriori.

4.6 Conclusion

Where does this bring us w.r.t. intensionality and meaning in mathematics? I think that what is clear is that the fact that *to have a genuine meaning* is *to have an intension* holds for mathematical expressions just as for empirical ones. However, to make sense of the concept of intension as applied to mathematical expressions we must disregard the almost universal acceptance of PWAI and slightly alter our understanding of it. I think that, roughly, an expression has an intension iff it does not only have an extension,[9] but if there are rules determining which changes of the extension are admissible so that it still counts as the same extension, and which changes would move us beyond this border.

Take (1*), i.e. the sentence *PA is consistent*: I do not think that anyone accepting it would be taken as not knowing what PA *means*. If you know the axioms of PA, but do not know about the consistency proofs, you would hardly be considered as not knowing what PA is, you would most probably be considered as merely missing something *about PA*. This seems to me to indicate that the rules of our language, by themselves, do not exclude the inconsistency of PA – at least they do not exclude it in the way in which the rules governing *not* exclude the truth of *Fido is a dog and Fido is not a dog*.

One of the morals to be drawn from this is that though the idea of epistemologically possible, but ontologically impossible, possible worlds is not wrong, we should not take possible worlds as the ultimate unexplained explainers. The ultimate level of explanation is the level of the rules of language. The sentence *Fido is a dog and is not a dog* is necessarily true because whoever would accept it would be deemed an incompetent speaker of English, in particular not knowing what *not* means.

Does this attitude towards meaning not lead us to an excessively fuzzy notion of meaning? If meaning is a matter of the rules of usage, the boundary of which can never be clearly specified, does it not mean that our theory of meaning goes awry? Not really – it is the meaning itself, not our theory of it that is to blame. We should become reconciled to the fact that meaning is not an object with sharp boundaries, but rather a role conferred on an expression by an often vague cluster of rules (as Wittgenstein, Quine, and Sellars all taught us.[10])

This, however, does not imply a humptydumptyism according to which expressions can mean whatever anybody wants them to mean. There are clear arguments distinguishing predicates that do mean *consistent* from co-extensional predicates that do not. They probably do not allow us to delimit the former ones quite sharply; hence they may not clear all the puzzles of the 'intensional' problems of mathematics, such as the Second Incompleteness Theorem. However, they do allow us to gain a much clearer view of their nature than when we ignore the intensionality of mathematical expressions or when we settle for its trivial theory.

Acknowledgements Work on this paper was supported by Grant No. 17-15645S of the Czech Science Foundation.

[9] To be sure, there may be expressions, especially definite descriptions, that have an intension but do not have an extension – in the case of mathematical expressions an example might be the greatest prime. But this is because they are composed out of meaningful subexpressions in such a way that they are themselves meaningful, though they do not pick up any extension.

[10] See Peregrin (2014).

References

Auerbach, D. 1985. Intensionality and the Gödel theorems. *Philosophical Studies* 48: 337–351.

Auerbach, D. 1992. How to say things with formalisms. In *Proof, logic and formalization*, ed. M. Detlefsen, 77–93. London: Routledge.

Bigelow, J.C. 1978. Believing in semantics. *Linguistics and Philosophy* 2: 101–144.

Boolos, G. 1995. *The logic of provability*. Cambridge: Cambridge University Press.

Carnap, R. 1947. *Meaning and necessity*. Chicago: University of Chicago Press.

Cresswell, M.J. 1975. Hyperintensional logic. *Studia Logica* 35: 25–38.

Detlefsen, M. 1986. *Hilbert's program: An essay on mathematical instrumentalism*. Dordrecht: Reidel.

Feferman, S. 1960. Arithmetization of metamathematics in a general setting. *Fundamenta Mathematicae* 49: 35–92.

Frege, G. 1892a. Über Sinn und bedeutung. *Zeitschrift für Philosophie und philosophische Kritik* 100: 25–50.

Frege, G. 1892b. Über Begriff und Gegenstand. *Vierteljahrschrift für wissentschaftliche Philosophie* 16: 192–205.

Halbach, V., and A. Visser. 2014a. Self-reference in arithmetic I. *The Review of Symbolic Logic* 7: 671–691.

Halbach, V., and A. Visser. 2014b. Self-reference in arithmetic II. *The Review of Symbolic Logic* 7: 692–712.

Hintikka, J. 1975. Impossible possible worlds vindicated. *Journal of Philosophical Logic* 4: 475–484.

Kripke, S. 1963a. Semantical considerations on modal logic. *Acta Philosophica Fennica* 16: 83–94.

Kripke, S. 1963b. Semantical analysis of modal logic I (normal modal propositional calculi). *Zeitschift für mathematische Logik und Grundlagen der Mathematik* 9: 67–96.

Kripke, S. 1965. Semantical analysis of modal logic II (non-normal modal propositional calculi). In *The theory of models*, eds. L. Henkin, J.W. Addison, and A. Tarski, 206–220. Amsterdam: North-Holland.

Kripke, S. 1972. Naming and necessity. In *Semantics of natural language*, eds. D. Davidson, and G. Harman, 253–355. Dordrecht: Reidel. Later published as a book.

Lewis, D. 1972. General semantics. In *Semantics of natural language*, ed. D. Davidson, and G. Harman, 169–218. Dordrecht: Reidel.

Löb, M. 1955. Solution of a problem of Leon Henkin. *The Journal of Symbolic Logic* 20: 115–118.

Montague, R. 1974. *Formal philosophy: Selected papers of R. Montague*. New Haven: Yale University Press.

Partee, B. 1982. Belief sentences and limits of semantics. In *Processes, beliefs and questions*, ed. S. Peters, and E. Saarinen, 87–106. Dordrecht: Reidel.

Peregrin, J. 1995. *Doing worlds with words*. Dordrecht: Kluwer.

Peregrin, J. 2006. Extensional vs. intensional logic. In *Philosophy of logic*, Handbook of the philosophy of science, vol. 5, ed. D. Jacquette, 831–860. Amsterdam: Elsevier.

Peregrin, J. 2014. *Inferentialism: Why rules matter*. Basingstoke: Palgrave.

Rosser, J.B. 1936. Extensions of some theorems of Gödel and Church. *Journal of Symbolic Logic* 1: 87–91.

peregrin@flu.cas.cz