

---

# A case for satisfaction classes: model theoretic vs axiomatic approaches to the notion of truth

ANDREA STROLLO

---

**ABSTRACT.** One of the basic question we can ask about truth in a formal setting is what, if anything, we gain when we have a truth predicate at disposal. For example, does the expressive power of a language change or does the proof strength of a theory increase?

Satisfaction classes are often described as complicated model theoretic constructions unable to give useful information toward the notion of truth from a general point of view. Their import is narrowed to a dimension of pure technical utility and curiosity. Here I offer an application of satisfaction classes in order to show that they can have a relevant role in confronting proof theoretical equivalent theories of truth.

*Keywords:* truth, satisfaction classes, axiomatic theories of truth, expansions, conservativity

## 1 Tarskian truth

The (broadly) Tarskian theory of truth has a prominent role in the field of formal truth theories, and it is the forced starting point for any further reflection toward the notion of arithmetical truth. Sometimes, such a theory is also called ‘*there is a full (not inductive) satisfaction class*’ or, shortly,  $PA(S)-$ . It consists, apart from the axioms of the base theory  $PA$  (Peano Arithmetic in its usual first order formulation), of the truth-compositional axioms inspired by the familiar Tarskian definition of truth. In other words  $PA(S)-$  is the theory in the language  $LTr := L_{PA} \cup \{Tr\}$ , yielded by the union of the axioms of  $PA$  in  $L_{PA}$  (which means that we do not have full induction) with the truth principles:

1.  $\forall\varphi(Atomic(\varphi) \rightarrow Tr(\varphi) \leftrightarrow (Tr^*(\varphi))$ ;
2.  $\forall\varphi(Tr(\neg\varphi) \leftrightarrow (\neg Tr(\varphi)))$ ;
3.  $\forall\varphi\forall\psi(Tr(\varphi \& \psi) \leftrightarrow (Tr(\varphi) \& Tr(\psi)))$ ;
4.  $\forall\varphi\forall i((Tr(\forall vi(\varphi))) \leftrightarrow \forall tTr(\varphi(t/vi)))$ <sup>1</sup>.

This is the usual way of writing down the axioms and, though comfortable, it is, strictly speaking, incorrect. In fact, a lot of coding apparatus has been suppressed to achieve a greater readability. To be rigorous we should write axiom 2, for example, like this:

$$\forall x\forall y\{Sent(x) \& Sent(y) \& Neg(y, x) \rightarrow [Tr(y) \leftrightarrow \neg Tr(x)]\}.$$

Here I shall persist with the most perspicuous presentation, but keep in mind the right form. The name  $PA(S)$ — should then be explained, because it carries important information.

When studying truth theories, it is often said that a background theory of syntax is needed. Without it, formulating axioms for a truth predicate and working out simple operations is impossible. We need to ascribe truth to so called ‘truth-bearers’, and a theory of syntax is intended to give us basic information about how these entities behave. One would expect a theory of syntax to consist of principles about linguistic expressions and this was exactly the case in the original work of Tarski. However, explicit formal theories of syntax, in the style of concatenation theories i.e., are not much widespread among truth-theorists. The reason is that, after Gödel, we know that a very good deal of syntax can be developed in  $PA$  (as in even weaker arithmetical theories) and, accordingly, we can correlate natural numbers and symbols of the language of  $PA$ . There are many ways to think of this correspondence between strings of symbols and numbers, but one often adopted is the easiest one: strings of symbols are identified with corresponding numbers.

---

<sup>1</sup>I am here assuming a substitutional interpretation of quantifiers.

## 2 Satisfaction classes and recursive saturation<sup>2</sup>

### 2.1 Non Standard Sentences and Satisfaction Classes

Thanks to Gödel's arithmetization of syntax we can use formulas of the language of  $PA$ ,  $L_{PA}$ , in order to talk about the syntax of this very language. In particular we find a correspondence between the set of sentences in  $L_{PA}$  and the elements of the domain of  $N$ , the *standard* model of arithmetic, whose domain contains all and only *standard* natural numbers.

One of the immediate consequences of compactness is that  $PA$  has, beside the standard model  $N$ , also different models, non isomorphic to  $N$ ; it has *non standard* models<sup>3</sup>. Let  $M$  be one of these non standard models, what would happen if we were to use  $M$  instead of  $N$  as a base for arithmetization? What would happen if we coded expressions of our language using not standard elements in the domain of  $N$  but those that are in the domain of  $M$ ? What would happen if we also used non standard numbers?

The first consequence would be the existence, beside *standard sentences* (those sentences coded by standard numbers), of new strange *non standard sentences*, coded by non standard elements in  $M$ <sup>4</sup>. In fact, among the many syntactical properties that can be represented in  $PA$  we can obviously define that of being a sentence of the language of  $PA$ , since there is a formula ' $Sent(n)$ ' which is true of  $n$  if and only if  $n$  is a code of a sentence of  $L_{PA}$ . Until we consider the standard model of  $PA$ , as is natural doing, this works as expected. However, in non-standard models, the formula ' $Sent(x)$ ' is going to be satisfied by non-standard numbers too. The reason is the *Overspill Principle*. According to it, if a formula is such that infinitely many *standard* numbers satisfy it, then — when we have a non-standard model — some *non-standard* number will satisfy it too. Clearly the formula ' $Sent(x)$ ' is satisfied by infinitely many

---

<sup>2</sup>The literature on satisfaction classes and recursive saturation is highly technical and difficult. We make general reference to [6], [2], [9], [8]. Personally, I have to thank Fredrik Engström for his patience in explaining me the quibbles of satisfaction classes. I certainly owe what I have understood (if any) to him and his long mails.

<sup>3</sup>For a good brief introduction to non standard models see [1, Ch. 25]

<sup>4</sup>The fundamental work is [13].

standard numbers, so when  $M \models PA$  is non standard, we have that  $M \models Sent(b)$  for some  $b \in M$ , and  $b$  non standard (similar for the syntactical notion of Term, Formula and so on). The existence of non-standard numbers that, according to the model, code sentences, drags us towards the realm of non-standard sentences. Very roughly, non standard sentences are sentences with a *non-standard structure*.

Non-standard sentences can be identified with non-standard elements that the model  $M$  ‘thinks’ to be actual sentences (those non-standard numbers that code sentences in the sense of  $M$ ). It is not easy to give a clear description of what non standard sentences are. I propose just an example. Consider the sentence in  $L_{PA}(\neg 0 = 0)$ . This is a case of a standard sentence that  $N$  (and then  $M$ ) recognizes to be a sentence, and such that it can be identified with its standard natural number of Gödel. We have a similar example with  $(\neg 0 = 0) \& (\neg 0 = 0)$  and  $(\neg 0 = 0) \& (\neg 0 = 0) \& (\neg 0 = 0)$ , where the number of conjuncts is a standard natural number, (2 and 3 respectively). If the number of conjuncts is a non-standard number, for instance  $(\neg 0 = 0) \& (\neg 0 = 0) \& \dots \& (\neg 0 = 0)$  (where the dots ‘...’ stand for  $a$ -many repetitions of the sentence  $(\neg 0 = 0)$ , and  $a$  is a non-standard number) we do not deal with a standard sentence anymore ( $N$  cannot recognize it as a sentence), we have obtained a non-standard sentence and  $M$  (if it contains  $a$ ) can recognize it to be a sentence.

Regarding non-standard sentences a natural question is whether and how they are true. We know that a truth predicate, ‘ $Tr$ ’ such that  $N \models Tr([\varphi]) \leftrightarrow \varphi$  for every sentence  $\varphi$  (where  $[\varphi]$  is the code of  $\varphi$ ), is not definable in  $L_{PA}$  (for Tarski’s undefinability theorem). This is the reason why we had to add such a new predicate together with axioms governing it, obtaining the truth theory  $PA(S)$ –. The same claim holds for non-standard sentences, since, a fortiori, a predicate ‘ $\Sigma$ ’ such that  $M \models \Sigma([\varphi]) \leftrightarrow \varphi$ , for every sentence  $\varphi$  in the sense of  $M$  (standard and non-standard) cannot be defined in  $M$ . What should such a predicate  $\Sigma$  be like? First of all it should agree, at least, with the truth predicate for standard sentences, namely it should respect Tarskian clauses. Here is where the notion of *satisfaction class* must be introduced.

Once we have added the truth predicate to PA and we have obtained our truth theory  $PA(S)-$ , we need to find an extension  $S$  in  $M$  for such a new predicate. Namely, given a model  $M \models PA$ ,  $S$  is supposed to be the set of numbers satisfying the axioms of  $PA(S)-$ ,  $M, S \models PA(S)-$ . When, in a model  $M \models PA$ , such a set  $S$  is available, we say that  $S$  is a *satisfaction class* for  $M$ . This explains the name ‘*PA plus there is a full (not inductive) satisfaction class*’. A satisfaction class<sup>5</sup>  $S$  over a model  $M$ , then, is a set  $S$  of elements in  $M$ , where any element  $b$  in  $S$  is such that  $M \models Sent(b)$  and  $b$  satisfies the Tarskian clauses for truth, namely the axioms of  $PA(S)-$ . In other words,  $S$  is a suitable extension for the truth predicate, as governed by Tarskian axioms, possibly including (codes of) non-standard sentences.

DEFINITION 1.<sup>6</sup> If  $M$  is a (non-standard) model of  $PA$ , a subset  $S$  of  $M$  is a *satisfaction class* if and only if:  $M, S \models PA(S)-$ .

Satisfaction classes can be classified further as follows.

DEFINITION 2. A satisfaction class  $S$  on  $M$  is *full* if, for every  $M \models Sent([\varphi])$ , we have that  $[\varphi] \in S$  or  $[\neg\varphi] \in S$ .

DEFINITION 3. A satisfaction class  $S$  on  $M$  is *partial* if and only if there is  $\alpha$  belonging to  $M \setminus N$  such that if  $M \models Sent[\varphi]$  and  $[\varphi] < \alpha$ , then  $[\varphi] \in S$  or  $[\neg\varphi] \in S$ .

The idea here is just that a satisfaction class is *full* if, for every formula  $\varphi$ ,  $S$  contains either  $\varphi$  or its negation and, if the satisfaction class is *partial* this is true only for those sentences coded by a (non-standard) number less than  $\alpha$ . Since standard sentences are coded by standard natural numbers and every standard natural number is less than every non-standard natural number, it follows that every satisfaction class (full or partial) behaves in the same way (they are full) with respect to standard sentences. It is important to notice that a satisfaction class, even if it is a partial one, has to decide some non-standard sentence, otherwise we have not a satisfaction class at all<sup>7</sup>.

---

<sup>5</sup>Though it is called ‘class’ it is a set.

<sup>6</sup>See [8] and [11].

<sup>7</sup>I owe this important remark to Fredrik Engström.

Indeed, if  $M$  is non standard we always have some non standard sentence into the extension of 'Tr'. Since every axiom of  $PA(S)-$  is subjected to a clause stating that the truth predicate applies to elements satisfying the formula ' $Sent(x)$ ', when we have a non-standard model, non-standard numbers can well enter into the range of the truth predicate. Actually this is not only possible but mandatory. In fact,  $PA(S)-$  proves  $\forall\varphi\{Sent(\varphi) \rightarrow [(Tr([\varphi]) \vee Tr([\neg\varphi]))]\}$ , thus, for every  $\varphi$  such that  $M \models Sent([\varphi])$  either  $\varphi$  or  $\neg\varphi$  must be in the extension of 'Tr', even if  $\varphi$  is non-standard.

DEFINITION 4. A satisfaction class is *inductive* if and only if the expanded structure  $(M, S)$  satisfies all the induction axioms for every formula in the language  $L_S = L_{PA} \cup \{S\}$  (Where the new symbol 'S' is governed by axioms stating that  $S$  is a satisfaction class. In our cases 'S' is substituted with 'Tr').

If this is the case, we have '*PA plus there is a full inductive satisfaction class*', turning from  $PA(S)-$  to  $PA(S)$ . These two theories have very different features and strength, but I shall mostly consider  $PA(S)-$  only.

Crossing these definitions we can get other classifications by distinguishing, with respect to full and partial satisfaction classes, those satisfaction classes that are inductive and those that are not.

We saw that, as far as standard sentences are concerned, a satisfaction class agrees with the traditional characterization of Tarskian truth. The surprising news is that classical compositional axioms are not enough to shape the truth of non-standard sentences. In other words, in order to characterize the truth (or falsity) of non-standard sentences we need some other tool than just compositional clauses in their standard formulation. If we only stick with Tarskian axioms, then we are free of constructing many different satisfaction classes such that they will agree on standard sentences but will disagree on many non-standard sentences. As a result, if a model  $M$  admits a satisfaction class, then it admits *many* satisfaction classes. Indeed, if  $S$  is a full satisfaction class for a countable  $M$ , then there are continuum many non isomorphic expansions  $(M, T)$  which are all elementarily equivalent to  $(M, S)$ .

This richness should be contrasted with this other fact: though a model can have many different satisfaction classes, not every model of  $PA$  can have one. Non *recursively saturated models*, in fact, cannot have any satisfaction class.

## 2.2 Recursive saturation

In order to explain the notion of *recursive saturation* we need some minimal elementary information<sup>8</sup>.

DEFINITION 5. If  $B$  is a theory, a *type* over  $B$  is:

- (i) a set  $P(\underline{x})$  of formulas containing a finite number of free variables  $\underline{x}$  (' $\underline{x}$ ' stands for a sequence of variables).
- (ii)  $P(\underline{x})$  is such that  $B \cup \{\varphi(\underline{c}) \mid \varphi(\underline{x}) \in P(\underline{x})\}$  is consistent. (Where ' $\underline{c}$ ' stands for a sequence of — possibly new — individual constants).

DEFINITION 6. A type  $P(\underline{x})$  is *complete* if and only if  $T \cup P(\underline{x})$  is a syntactical complete theory (that is for every  $\varphi(\underline{x})$ ,  $T \cup P(\underline{x}) \vdash \varphi(\underline{x})$  or  $T \cup P(\underline{x}) \vdash \neg\varphi(\underline{x})$ ).

DEFINITION 7. A type  $P(\underline{x})$  is *principal* if and only if there is a single formula  $\psi(\underline{x})$  such that  $T \vdash \forall x(\psi(\underline{x}) \rightarrow \varphi(\underline{x}))$ , for every  $\varphi(\underline{x}) \in P(\underline{x})$ .

DEFINITION 8. If  $M \models B$ , a type  $P(\underline{x})$  is *realized* in  $M$  if and only if there is  $\underline{a} \in M$ , such that  $M \models \varphi(\underline{a})$ , for every  $\varphi(\underline{x}) \in P(\underline{x})$ . Otherwise  $M$  *omits* the type  $P(\underline{x})$ .

For completeness theorem, if  $P(\underline{x})$  is a type over a theory  $B$ , then  $B$  has a model that realizes  $P(\underline{x})$ . Similarly, if  $P'(\underline{x}), P''(\underline{x}) \dots$  are types over the theory  $Th(M)$  of a model  $M$  (that is the set of all the sentences  $\varphi$  such that  $M \models \varphi$ ), then there is an elementary extension  $M'$  of  $M$  that realizes every type  $P(\underline{x})$ .

DEFINITION 9. A type  $P(\underline{x})$  is *recursive* if the set  $\{\varphi(\underline{x}) \mid \varphi(\underline{x}) \in P(\underline{x})\}$  is recursive. (Notice that what is recursive is the set of formulas, not the formulas, which can have whatever complexity.)

---

<sup>8</sup>It is possible to give the following definitions also in model-theoretic terms instead of talking of theories. I use the proof theoretic definition in order to stress the relation of this notion with the notion of truth as axiomatized by  $PA(S)-$ .

DEFINITION 10. A model  $M$  is *recursively saturated* if and only if every recursive type over  $Th(M)$  is realized in  $M$ .

A recursively saturated model can be thought as a *big* and *homogeneous* model. A non-recursively saturated model  $O$  is a model where at least one recursive type (a recursive set  $P(\underline{x})$  of formulas) is not realized in  $O$ . This means that there are formulas  $\varphi(\underline{x}) \in P(\underline{x})$ , for which elements  $\underline{a}$  in  $O$  such that  $O \models \varphi(\underline{a})$  are not available. This can happen, for example, when the model is not *homogeneous* or it is not *big* enough. To make it such, we need to extend  $O$  to  $O'$  adding new elements  $\underline{a}$  with the desired features.

The fundamental fact now is that if a non-standard model admits a satisfaction class, then such a model needs to be big and homogeneous in this sense: it must be recursively saturated. This is the sense of Lachlan's theorem:

THEOREM 1 (LACHLAN'S THEOREM<sup>9</sup>). *If  $M$  is a non-standard model of PA with a full not inductive satisfaction class, then  $M$  is recursively saturated.*

It is possible to get a similar result also for partial satisfaction classes too:

THEOREM 2.<sup>10</sup> *If  $M$  is a non-standard model of PA with a partial satisfaction class, then  $M$  is recursively saturated.*

We can sum up the story stating that if  $M$  is non-standard and it has a satisfaction class (it does not matter whether full or partial, or whether it is inductive or not), then  $M$  must be recursively saturated. Recursive saturation is a necessary condition for a non-standard model to have a satisfaction class<sup>11</sup>.

However, recursive saturation is not a sufficient condition to guarantee the possibility of a satisfaction class: in fact there exist non countable recursively saturated models without a full satisfaction class or an inductive satisfaction class<sup>12</sup>. Recursive saturation of a

---

<sup>9</sup>[12]

<sup>10</sup>See [6, Theorem 15.5 and proposition 15.4].

<sup>11</sup>Notice that this is not true for the standard model  $\mathbb{N}$ .  $\mathbb{N}$  is not recursively saturated but it does admit a 'satisfaction class'.

<sup>12</sup>See [5].

non-standard model is a sufficient condition to have a satisfaction class only together with countability:

**THEOREM 3.** *If  $M$  is a countable recursively saturated model of  $PA$ , then  $M$  admits a satisfaction class.*

Notice that this does not mean that every countable recursively saturated model of  $PA$  admits a satisfaction class whatsoever: for instance it is not enough to have a full *inductive* satisfaction class. In fact,  $PA(S)$  is famously able to prove the consistency of  $PA$ , so that just models where  $Con(PA)$  holds can be expanded to models of  $PA(S)$ , and some recursively saturated models are excluded.

It is very important for us to stress that **there are non-standard models of  $PA$  such that they are not recursively saturated**. Therefore, there are non-standard models of  $PA$  that do not admit a satisfaction class.

### 3 Satisfaction classes and axiomatic truth theories

With such results available we can draw some important consequences. The first observation is rather natural and concerns the relation between satisfaction classes and axiomatic theories of truth. The notion of satisfaction class has been constructed with the purpose of characterizing the set of all arithmetical truths in a certain model from a model-theoretic point of view, while an axiomatic setting tries to characterize the behaviour of the truth predicate. It is clear that such approaches can be considered, in a certain measure, as two different ways of working at the same problem. We can then expect an axiomatic theory of truth to give an axiomatization of a predicate defining a satisfaction class<sup>13</sup>.

The second, more important observation, is that a model-theoretic approach to truth can enlighten aspects that a pure proof theoretic investigation is not able to show.  $PA(S)$ – gives us the clearest example of such a situation. In fact, from a proof theoretic

---

<sup>13</sup>In our definition of satisfaction class we used a relation symbol to talk about the satisfaction of a formula by a sequence of objects, while in the axiomatic theories we are using a one-place truth predicate. This difference, however, has not deep effects, at least with regard to our problems. It would have been possible, for example, to define a satisfaction class avoiding the notion of satisfaction (as in [2]).

point of view,  $PA(S)-$  is a *conservative* extension of  $PA$ . So with the addition of the truth axioms in  $PA(S)-$ , apparently, we do not get any new arithmetical information, and, viewed from  $PA$ , this enrichment looks rather useless. We might be tempted to say that  $PA$  and  $PA(S)-$  have the same arithmetical content. This, however, would be a mistake as the application of satisfaction classes will show.

To better explain the point, consider the theory  $ACA-^{14}$ , which is the axiomatic theory for second order arithmetic (*Arithmetic with Comprehension Axioms*), yielded adding to  $PA$  axioms for second order comprehension preventing full induction. More precisely:  $ACA-$  is formulated in the second order language  $L2$  of second order arithmetic and is given by adding to  $PA$  the axioms

$$\exists X \forall y (y \in X \leftrightarrow \varphi)$$

where  $\varphi$  is a formula of  $L2$  without any second order quantifier or  $X$ . Notice that in  $ACA-$  there is arithmetical induction only.

As is well known  $ACA-$  is conservative over  $PA$ . Moreover  $ACA-$  can define a truth predicate for  $PA$ . Namely, in  $ACA-$  it can be defined a formula  $\tau(x)$  such that  $ACA- \vdash \tau([\varphi]) \leftrightarrow \varphi$ . (Though a truth predicate respecting Tarskian clauses cannot be defined in  $ACA-$ .) Again a natural expectation might be that  $ACA-$  and  $PA(S)-$  should be two arithmetically equivalent theories, sharing the same *arithmetical content*. This is the natural conclusion given the fact that both additions are conservative over  $PA$ , so that they prove the same arithmetical theorems. Indeed, if we have full induction, turning to  $ACA$  and  $PA(S)^{15}$ , we obtain an interdefinability result. As is well known,  $ACA$  can define the truth predicate of  $PA(S)$ , and  $PA(S)$  can define membership of  $ACA$ . Thus, the moment we drop full induction, even if we lose this pleasant interdefinability, we might expect the equivalence to keep holding. As a matter of fact, however, these two theories are not arithmetically equivalent and the proposed proof theoretic analysis did not give us all relevant information. If we turn to a model

---

<sup>14</sup>See [4].

<sup>15</sup>These theories are exactly the same as  $ACA-$  and  $PA(S)-$  except from the fact that induction of  $PA$  is now extended to these new languages.

theoretic approach, instead, we see clearly that they have a deeply different arithmetical impact.

$ACA-$  is conservative over  $PA$  and each model of  $PA$  can be expanded to a model of  $ACA-$ . This is actually how the conservativity of  $ACA-$  is usually proved. So, in this sense,  $ACA-$  can be thought to really have the same arithmetical content of  $PA$ .  $ACA-$  keeps characterizing all the models of  $PA$ . Hence, if we see  $PA$  as a way to talk about a large class of (arithmetical) structures,  $ACA-$  can be seen as another way to talk of the very same class. This is a strengthening of the conservativity result: not only do not we get new arithmetical theorems, we do not restrict the number of possible models either.

From the supposed equivalence between  $ACA-$  and  $PA(S)-$  one would expect a similar situation to keep holding even for  $PA(S)-$ , which, in fact, is conservative over  $PA$  too. But, as we know, things are very different. Indeed, a non-standard model  $M$  of  $PA$  is a model of  $PA(S)-$  if and only if  $M$  admits a satisfaction class. Unfortunately, not every model of  $PA$  does. If a non-standard model of  $PA$  is not recursively saturated, then it is no use trying to expand it to  $PA(S)-$ . The reason is just Lachlan's theorem.

There is a kind of asymmetry between  $ACA-$  and  $PA(S)-$  then. Both are conservative over  $PA$ , but they affect the models of  $PA$  in very different ways.  $PA(S)-$  cannot be taken to be another way of characterizing the same class of structures since it is able to exclude some of the models of  $PA$ . Thus, claiming that  $PA(S)-$  has the same arithmetical content of  $PA$  is not plainly correct. It has extra-content, for it requires the models to be in a certain precise way: they must be recursively saturated.

#### 4 A philosophical application

A philosophical insight into the notion of truth is deeply related to similar issues. Consider, for example, a deflationist approach to truth, which is probably the most debated proposal nowadays. Deflationism rejects the traditional philosophical explanation of the concept. In particular, according to it, truth is claimed not to be a very deep notion, and the property it stands for, if any, not to have any metaphysical structure or weight. Opposite to traditional

approaches, like correspondence theories, truth is claimed to lack any robust ontological import: truth is an unsubstantial property.

Such an unsubstantiality has been explained by exploiting the fact that, if truth is characterized using some very simple axioms, exactly like typical Tarskian ones, we can easily get conservative extensions of  $PA$  in a number of cases. Since, then, conservative theories are thought to be innocent additions, by using the notion of conservativity it is possible to make sense of the metaphysical innocence of truth. If a theory  $T$  is conservative over a theory  $B$ ,  $T$  does not prove anything new regarding what  $B$  is about, so that, it has been suggested, viewed from  $B$ , the addition of  $T$  is redundant. If  $T$  is our theory of truth, then, truth is harmless in this sense, as deflationists argue. Consequently, the unsubstantiality of truth can be identified with the conservativity of its theory. If a truth theory is conservative, it will not prove anything apart from semantical claims, and it will not be capable of concrete (read extra semantical) power. The conservativity of a theory of truth can be taken to be evidence for the evanescent nature of the property of truth.

Things, however, are not so easy. In fact, as our previous analysis showed, there is much more content in a theory than that that can be enlightened by merely proof theoretical means. Possibly, a theory  $T$  can be conservative over  $B$ , while not every model of  $B$  be expandable to a model of  $T \cup B$ , so that  $T$  can still have some impact over what  $B$  is about. Conservative theories can as well affect the content of the theory they are added to. A truth theory like  $PA(S)-$  is conservative over  $PA$ , nevertheless it makes a difference.  $PA(S)-$  has a rich and interesting arithmetical content which is at least as rich as the notion of recursive saturation, a very pivotal tool in model theory. It follows that conservativity is not a suitable candidate to analyse the alleged unsubstantiality and innocence of truth; exactly as a mere proof theoretic approach is not enough to make us able to draw all the relevant consequences.

## 5 Conclusion

When we take an axiomatic approach to truth, the customary way of proceeding is that of devising a suitable set of axioms governing the truth predicate and then studying the logical properties of such

an axiomatization. In particular, a certain axiomatic theory of truth is evaluated with respect to its impact over an arithmetical theory like  $PA$ . We check whether it is conservative, whether it is able to prove new arithmetical theorems or whether it is able to prove the consistency of  $PA$  and so on. Once we have that, it is interesting to assess what the proof strength of our theory is with respect to other proposals on the market. We try to put our new axiomatization in the right place of the hierarchy.

The reflections I proposed are mainly motivational ones. They are intended to show that a merely proof theoretical evaluations of theories, and of truth theories in particular, are not, sometimes, fine grained enough. There might be more differences than those emerging in proof theoretical terms: two theories with the same arithmetical proof strength can have very different impacts over the model of  $PA$ , and exhibit a different arithmetical content. This is absolutely relevant if we are interested in identifying the correct power of truth theories. But this is also important from a general philosophical point of view, since, if this point is neglected, metaphysical misconceptions loom.

The notion of truth is traditionally tackled, into a formal framework, by building either axiomatic systems or semantical interpretations. Where, in the latter case, one proposes a model in which the truth predicate is interpreted in some nice way. Semantical and axiomatic approaches do not interact many often, so that the connection between them is rather weak. This is certainly unpleasant from a general point of view: a model theoretical approach can be crucial to enlighten important aspects of axiomatic theories and can offer an important contribution to their evaluation. The considerations above showed a case of this possibility and can be generally considered as a first step toward a closer collaboration between an axiomatic and a semantical approach to the notion of truth.

## References

- [1] BOLOS, G. S., BURGESS, J. P., and R. C. JEFFREY, *Computability and Logic*, Cambridge University Press, 2007.
- [2] ENGSTRÖM, F., *Satisfaction Classes in Nonstandard Models of First-order Arithmetic*, Thesis for the Degree of Licentiate of Philosophy,

- Department of Mathematics Chalmers University of Technology and Goteborg University, 2002.
- [3] FRIEDMAN, H. and SHEARD, M., An Axiomatic Approach to Self-Referential Truth, *Annals of Pure and Applied Logic* 33:1–21, 1987.
  - [4] HALBACH, V., *Axiomatic Theories of Truth*, Cambridge University Press, 2010.
  - [5] KAUFMANN, M., A Rather Classless Model, *Proceedings of the American Mathematical Society* 62(2):330–333, 1977.
  - [6] KAYE, R., *Models of Peano Arithmetic*, Oxford Logic Guides, Oxford University Press, 1991.
  - [7] KETLAND, J., Deflationism and Tarski's paradise, *Mind* 108(429):69–94, 1999.
  - [8] KOSSAK, R., A Note on Satisfaction classes, *Notre Dame Journal of Formal Logic* 26(1):1–8, 1985.
  - [9] KOTLARSKI, H., Full Satisfaction Classes: A Survey, *Notre Dame Journal of Formal Logic* 32(4):573–579, 1991.
  - [10] KOTLARSKI, H., KRAJEWSKI, S. and LACHLAN, A. Construction of Satisfaction Classes for Nonstandard Models, *Canadian Mathematical Bulletin* 24:283–93, 1981.
  - [11] KRAJEWSKI, S., *Nonstandard satisfaction classes*, Springer Lecture Notes in Mathematics, 537, 1976.
  - [12] LACHLAN, A. Full Satisfaction Classes and Recursive Saturation, *Canadian Mathematical Bulletin* 24:295–297, 1981.
  - [13] ROBINSON, A., On languages which are based on non-standard arithmetic, *Nagoya Mathematical Journal* 22:83–117, 1963.
  - [14] SHAPIRO, S., Proof and Truth: Through Thick and Thin, *Journal of Philosophy* 95:493–521, 1998.
  - [15] SHEARD, M., A Guide to the Truth Predicate in the Modern Era, *Journal of Symbolic Logic* 59(3):1032–1054, 1994.
  - [16] TARSKI, A., The Concept of Truth in Formalized Languages, in A. Tarski (ed.), *Logic, Semantics, Metamathematics*, Oxford: Oxford University Press, 1956, pp. 152–278.