A teaching room in an Oxbridge college. Fred Philonous, a tutor of the college, sits in front of a large desk. At the other end of the room two armchairs and a sofa are grouped around a fragile looking small table. In the background, close to the window, is a large whiteboard. Through the window perpendicular buildings can be seen. FRED is typing slowly on his computer. After a short time a student, Sophia, peeks in through one of the window panes. She disappears again and after a few seconds a knocking can be heard.

FRED Come in.

SOPHIA (entering the room.) Hi.

FRED (While speaking, he gets up from his chair and walks slowly to an armchair.) Hello. I got your essay. It was on paradoxes. I thought we had agreed that you write an essay on 'Should we define truth in terms of correspondence or coherence?'

SOPHIA No, you just said I should write an essay on truth.

FRED But the question on the correspondence and coherence theory was the next question on the reading list.

SOPHIA Yes, I skipped it. The next question was also on truth.

FRED The week before we talked about definitions of knowledge and the Gettier problem. There is a reason why I put the question on the correspondence theory as the next question on the list. In epistemology we use a correspondence definition of truth. Almost all epistemologists subscribe to a correspondence theory of truth and define truth in terms of correspondence with reality. It would have made sense to write on the correspondence theory because it would have been an analysis of the truth condition in definitions of knowledge.

*In this play I have used some motives from (Halbach and Visser 2014a) and (Halbach and Visser 2014b). My thanks go to all those who helped us with these two papers. Of course the present paper is my very personal take on some of the topics touched upon in the two joint papers.
SOPHIA I started reading Russell. But then they said in a survey article that Tarski showed that truth cannot be defined, and I thought that there is a general problem for definitions of truth in terms of correspondence or whatever, if Tarski is right.

FRED Yes, there may be a problem, but believe me: In an exam it’s much easier to write an essay on the correspondence or coherence theory of truth than on Tarski’s account of truth.

SOPHIA I am sorry, I answered the question ‘What is the source of semantic paradox?’

FRED (sinking slowly into the armchair.) Ok, so what is the root of evil?

SOPHIA (sitting down on the sofa) Some people say that self-reference is the source of all semantic paradoxes. Others say that it’s the truth predicate that can apply to sentences containing the very same truth predicate.

FRED Well, clearly, if we distinguish between object and metalanguage, the paradoxes disappear. That’s how mathematical logicians solve the liar paradox. But distinguishing between object and metalanguage doesn’t really help to solve the liar paradox in natural language. In natural language we don’t observe this distinction; we apply the truth predicate to all sentences whether they contain the truth predicate or not.

SOPHIA Even that isn’t clear to me. There’s this paradox by Visser.¹ It seems to show that we can still get into trouble even when we have only typed truth predicates. He has a hierarchy of languages as Tarski did; but he still gets a paradox.

FRED I think you are confused. Many students here become confused about these logical matters because they learn logic from a confusing textbook. (He gets up and paces slowly up and down) If you have truth predicates each of which apply only to sentences containing only truth predicates lower in the hierarchy, then all paradoxes are blocked. The liar paradox relies on a truth predicate that is applied to a sentence containing the same truth predicate. (He looks out of the window and continues after a pause.) Or maybe you have two truth predicates true₁ and true₂ and two sentences. Now you write on one side of a postcard ‘The sentence on the other side is not true₁’ and on the other side ‘The sentence on the other side is true₂’. When you think about it, the truth predicate ‘true₁’ is only applied to the sentence on the other side and this sentence contains only the truth predicate ‘true₁’. Conversely, the sentence on the other page says that the sentence on the first page is true₂. So, the sentence on the second page is only applied to a sentence containing ‘true₁’ but not ‘true₂’ itself. So none of the sentences talks about a sentence that contains the same truth predicate. But still, both sentences are self-referential. Both sentences refer to themselves. Their self-reference is indirect. The source of paradox is self-reference. (He stops walking behind his armchair.)

SOPHIA I have come across the postcard paradox. But I mean you can get a paradox without any circularity of this kind.

FRED Um, I remember. Yablo’s paradox is supposed to show that there is semantic

¹See (Visser 1989). Drafts of this paper were circulated in the early 1980s.
paradox without self-reference.²

SOPHIA (interrupting FRED) No, I don’t mean Yablo’s paradox. Yablo used a truth predicate that is applied to sentences that contain the same truth predicate. In Visser’s paradox there are many truth predicates linearly ordered in a hierarchy. Each of these predicates applies only to sentences with predicates that are strictly lower in the hierarchy.

FRED But didn’t Tarski prove that such a hierarchy blocks the paradox?

SOPHIA The weird feature of Visser’s hierarchy is that it is illfounded.

FRED What does that mean?

SOPHIA We have truth predicates ‘trueₙ’ for every natural number ₙ. For a given ₙ, ‘trueₙ’ is a truth predicate for sentences that contain only truth predicates with an index bigger than ₙ. This means that the predicate ‘true₀’ is the topmost predicate; it applies to sentences possibly containing ‘true₀’, ‘true₁’ and so on. The predicate ‘true₁’ is one level lower in the hierarchy; it applies to sentences possibly containing ‘true₂’ ‘true₃’ and so on, but not to sentences containing ‘true₀’ or ‘true₁’.

FRED (sinking slowly back into his armchair) Sounds a little weird. But it’s still a hierarchy. What’s paradoxical about it?

SOPHIA (jumps up from the sofa and walks to the whiteboard.) Unlike Yablo, Visser used the Gödel fixed-point theorem for defining the problematic sentences, but we can write them just as a list, in the way Yablo did. (After trying out some pens, Sophia finds one with some ink left in it and writes the following lines on the board:)

No sentence below this one is true₀.
No sentence below this one is true₁.
No sentence below this one is true₂.
No sentence below this one is true₃.
⋮

This list shows that we can have paradox without a global truth. The truth predicates apply only to sentences containing truth predicates lower in the hierarchy.

FRED (interrupting) What? If we assume the first sentence at the top, the sentence in the second line isn’t true₀. Therefore some sentence below the second line must be true₁, thereby contradicting the first line. So, the first line can’t hold.

SOPHIA Exactly! Therefore there must be sentence below the topmost sentence that is true₀.

FRED (sounding bored) But assuming one of the sentence lower down leads to a contradiction in the same way that assuming the first line led to a contradiction. That looks very much like Yablo’s paradox.

SOPHIA (erasing the type indices with a very dirty sponge) Yes, you get Yablo’s list by erasing the indices. (This leaves the following result on the whiteboard:)

²See (Yablo 1985, p. 340), (Yablo 1993) and the recent monograph (Cook 2014).
FRED  So, why did you bring up Visser’s paradox at all?

SOPHIA  Because you claimed that the semantic paradoxes disappear if we apply the object/metalanguage distinction. I think Visser’s paradox shows that you can get a paradox even if you observe the distinction and you admit only truth predicates that apply to sentences with truth predicates lower in the hierarchy. Yablo’s paradox, in contrast, is merely another proof of the inconsistency of the full T-sentences. We knew that before. But we didn’t know that even the hierarchical approach can lead to paradox. After all, it was Kripke who mentioned the illfounded hierarchies.³ There is no hint that Kripke realized that there was a paradox. Let me see. (She takes her computer out of her back, opens and continues after searching on the computer.) Here. Kripke writes: ‘Even if unrestricted truth definitions are in question, standard theorems easily allow us to construct a descending chain of first-order languages \(L_0, L_1, L_2, \ldots\), such that \(L_i\) contains a truth predicate for \(L_{i+1}\).’ That doesn’t sound as if he had been aware that such a hierarchy doesn’t have a standard model.

FRED  Ok, so Visser gave a proof of a new result, and Yablo gave a new proof of an old result.

SOPHIA  Yes, and Visser showed that strict hierarchies of truth predicates don’t save us from paradox. He showed that we can have paradox without such a global predicate. For his paradox we use truth predicates that apply only to sentences containing truth predicates lower in the hierarchy. In this sense Visser’s paradox is much stronger than Yablo’s paradox.

FRED  Well, which axioms does Visser need to get the inconsistency? What does Yablo need?

SOPHIA  (sitting down again) There are different ways to formalize these paradoxes. But usually you get only an \(\omega\)-inconsistency. For Yablo’s paradox you need something like the unrestricted T-sentences. For Visser’s paradox you need the T-sentences for instances of lower levels. Thus his T-sentences obey Tarski’s object/metalanguage distinction.⁴

FRED  (staring out of the window) Ok, then the problem isn't the unrestricted truth predicate. The root of the evil must be self-reference. Hasn't Priest shown that a proper definition of the Yablo list needs the fixed-point lemma or the recursion theorem or something that involves self-reference?⁵ Doesn’t Visser need that too?

SOPHIA  Yes, Visser used the fixed-point lemma. When you formulate all this in arith-

³See (Kripke 1975, p. 697).
⁴Cf. (Leitgeb 2001) for a comparison of Yablo’s with Visser’s and other paradoxes.
⁵See (Priest 1994).
metic, you would normally use the fixed-point lemma or the recursion theorem.

FRED Why ‘normally’?

SOPHIA Priest and others pointed at the use of the fixed-point lemma in the proofs of the inconsistency or $\omega$-inconsistency. You can also use the recursion theorem. That’s even nicer because you define a function that corresponds to the sentences in the Yablo or Visser lists. But I don’t know a proof that the use of the fixed-point lemma or the recursion theorem cannot be avoided and other proofs cannot be given. It don’t even know what exactly it means to say that Yablo’s or Visser’s list cannot be defined without the fixed-point lemma or the recursion theorem. You could always change the formulation of the fixed-point lemma a little to show that you don’t have to go exactly through a certain lemma.

FRED Yes, there are variants, but in the end it all comes down to the old Gödel fixed point lemma or a trivial variation thereof. And that means in the end that the root of paradox is self-reference. Gödel himself already described his sentence as a sentence that states its own unprovability. Gödel produced a sentence that is equivalent to the sentence saying that the sentence isn’t provable. (He gets up and writes in the corner of the whiteboard:)

\[ G \iff A('G') \]

If this equivalence is provable, then $A$ says about itself that it has the property expressed by the formula $A(x)$, whether $A(x)$ expresses non-provability, falsity, or still something else.

SOPHIA Really? Truth predicates are supposed to have many fixed points. Probably you want

\[ G \iff \text{true('G')} \]

for many sentences, not just for sentences $G$ that are constructed using Gödel’s trick? For instance, you want to have it for ‘snow is white’ as an instance. ‘Snow is white’ is a fixed point of the truth predicate. But do you really want to claim that it’s self-referential?

FRED (sinking into his armchair with a sigh) Err, no. To be self-referential, the sentence must have been obtained via the Gödel’s construction. And vice versa, if a sentence has been obtained via that construction, it’s self-referential.

SOPHIA That’s strange. I read that Gödel’s diagonal construction is a trick. Now you seem to claim that it’s an analysis of self-reference and the only way to get self-reference.

From time to time some people pass the window. While FRED is speaking, a lady passes with two plastic bags full of scripts. One of the scripts slips out, hovers for a second in the air and then sinks slowly down to the floor.

FRED Well, in arithmetic Gödel’s construction is the only way to get self-reference. In English there are more options. Think of the notorious sentence ‘I am not true’. It sounds a little odd, because the personal pronoun usually refers to the speaker, not
to the sentence that is uttered. We can also get self-reference via description. For instance, we can express the liar by saying: (whispering) The whispered sentence isn’t true.

SOPHIA Sentences with a first-order personal pronoun are used when people make claims about themselves. People usually don’t use descriptions to make claims about themselves. It would be very strange if I referred to myself by a definite description. I say ‘I will write an essay on truth for next week’ and not ‘The person sitting on the sofa will write an essay on truth for next week.’ Personal pronouns seem to be the most direct way to make claims about oneself. Of course pretending that a sentence can make claims about itself using a first-person personal pronoun is weird. But this entire way of talking about making claims about itself and so on is just a metaphor that treats sentences like persons that can make claims about others and themselves.

The lady with the plastic bag has returned and can be seen through the window again. She picks up the lost scripts and continues her way again, while she is being watched through the window by FRED and SOPHIA.

FRED Why?

SOPHIA Assume that a few seconds ago the woman spotted the script on the floor. When she thought ‘I lost that piece of paper,’ she had a belief about herself. If she had just thought that the last person who passed the window lost the script, without realizing that it was her, she wouldn’t have come back. She thought that she herself lost the script.

FRED Even if she had thought that the last person who passed the window lost the script, it would still be a belief about herself, because she was that last person. Using the personal pronoun is just shorter than a definite description.

SOPHIA Is it really a belief about herself in a strict sense? The belief ‘The last person who passed by lost the script’ is only accidentally about herself, and she needn’t know that it’s about herself. If she thinks ‘I lost that script,’ then it’s not accidental – and only the latter belief prompts her to pick it up again.

FRED Ok, by using the pronoun ‘I’ she would refer to herself without relying on any description. She would have a de re believe about herself that she lost it, as far as it ok to say that she is a res. But in arithmetic these differences play no role whatsoever. ‘The last person who passed the window’ refers to whoever passed the window, while the first-person pronoun ‘I’ rigidly always refers to the speaker. You remember what we said when we talked about Naming and Necessity?6

SOPHIA Of course.

FRED In arithmetic the differences don’t matter, because in arithmetic nothing is contingent or a posteriori. Whatever satisfies a definite description in arithmetic does so by necessity and a priori. So all designators are rigid.

6Fred refers to (Kripke 1972).
SOPHIA What I have in mind isn’t the *de re/de dicto* distinction. I mean something that is stronger than *de re*. It’s more what Lewis called *de se*. The lady could think: ‘Somebody must have been there, in front of the window. That person must have lost the script.’ She would believe something about this person in a *de re* way, referring to him or her using the phrase ‘that person’; but she need not believe *de se* that she herself lost the paper.

FRED I think that you confuse two topics here. We don’t talk about propositional attitudes here. That’s a topic for a later week.

SOPHIA (getting up again) Perhaps this really hasn’t anything to do with the liar paradox and self-reference in logic. But *de se* beliefs are the paradigmatic self-referential beliefs.

FRED Why does this matter?

SOPHIA (wiping the whiteboard again) Because whether a sentence is true or not can depend on how the sentence refers to itself when we don’t use pronouns like ‘I’. I am not sure.

FRED The main point is that we have a term in the language that refers to the sentence itself – or its code.

SOPHIA Consider the following sentence (writing on the whiteboard and reading out the following sentence:)

\[(A) (A) = (A)\]

This sentence says about \((A)\), which happens to be the very sentence, that it’s identical to \((A)\). You would say that the sentence says about itself that it’s identical to \((A)\), don’t you?

FRED Of course.

SOPHIA Now consider another sentence (again writing on the whiteboard and reading out:)

\[(B) (B) = (A)\]

This sentence says the same about itself, namely that it is identical to \((A)\). They say the same thing about themselves, according to you. Both sentences claim about themselves that they are identical to \((A)\). Although they say the same thing about themselves, \((A)\) is true while \((B)\) is false.

FRED Actually I retract my claim that \((A)\) says about itself that it is identical with \((A)\). It says about itself that it is self-identical. I mean, \((A)\) states that it is identical with itself. If we allow sentences to refer to themselves in the first person, then the sentence \((A)\) expresses the same proposition as the sentence ‘I am identical with myself.’

\footnote{See (Lewis 1979).}
SOPHIA Now you resort to the pronouns used to express de se-beliefs. But neither (A) nor (B) contain a personal pronoun. They don’t express anything de se. You understand the label ‘(A)’ as a first-person pronoun. The sentence ‘I am identical to myself’ says only one thing of itself. It only says that it is self-identical. But as soon as you use some other designator such as the name ‘(A)’, that is no longer the case. Our sentence (A) says about itself that it is self-identical, but also that it is identical to (A) and that (A) is identical to it. I could have used another example with a relation that isn’t symmetric. I could have written (writing and speaking the following sentence):

(C) (C) comes earlier in the alphabetical order than (C)

By ‘coming earlier in the alphabet’ I mean ‘coming earlier if we ordered all sentences in the usual alphabetical order’. Sentence (C) says at least three things about itself; or, if you prefer, it ascribes three properties to itself. First, it says that it comes earlier in the alphabetical order than itself. But it also says about itself that (C) comes earlier than it – and that it comes earlier than (C). These are three different things. We can also construct sentences that ascribe to themselves only one of those two last properties: (continues writing):

(D) (C) comes earlier in the alphabetical order than (D)

and also

(E) (E) comes earlier in the alphabetical order than (C)

The sentence (C) is false, (D) is true, because C precedes D in the alphabet; (E) is false again.

FRED That’s confusing.

SOPHIA All the confusion disappears once you use first-order pronouns.

FRED I think they add extra complications.

SOPHIA Actually they simplify the situation. The sentence ‘I am identical to myself’ ascribes to itself only the property of self-identity; it only says that it is identical to itself. You read (A) in this way (pointing at (A)). You think that (A) says the same as ‘I am identical to myself’. But (A) ascribes at least two properties to itself, while the sentence ‘I am identical to myself’ ascribes only one property to itself.

FRED (frowning) What?

SOPHIA We have the sentence ‘I am identical to myself’. It says about itself that it is self-identical. And we have (A). That sentence makes probably three claims about itself. That’s why first-person pronouns and reflexive pronouns simplify everything. When we say that a sentence says something about itself, e.g., that it’s not true, not provable, true, self-identical, and so on, we use the reflexive pronoun ‘itself’ in the same way that we said about the woman that she thought about herself that she had lost that script.
FRED Sorry, how is that related to her?

SOPHIA If the woman says 'I lost that script', she ascribes to herself the property of having lost that script. It's the same with the sentence 'I am identical to myself'. It ascribes to itself the property of being self-identical.

FRED Sure.

SOPHIA If the woman thinks 'The last person to pass by lost the script', then it's far less clear whether she ascribes any property to herself. Probably she does ascribe to herself a property, but that's contingent on the assumption that she is the last person to pass the window. She doesn't ascribe to herself the property \emph{de se}. It's only second-rate self-reference. First-rate self reference is obtained by using first-person pronouns.

FRED How is all this relevant for self-reference in formal languages where we get self-reference via the diagonal lemma?

SOPHIA We describe the Gödel sentence as the sentence as the sentence that says: 'I am not provable', which we take to mean that the sentence says of itself that it's not provable. Similarly, we describe the liar as the sentence that says: 'I am not true.' And we describe the truth teller as the sentence that says: 'I am true.' Then we study these sentences in formal languages that don't have personal pronouns or a similar device. In these languages we have only definite descriptions. We study sentences corresponding to the claim 'The last person who passed the window lost the script.' We don't study sentences of the form 'I lost the script.' We don't have first-person pronouns in arithmetic. There we have only definite descriptions of the sort 'the result of substituting the numeral of ...'. The allegedly self-referential sentences in formal languages such as the language of arithmetic are not \emph{de se}.

FRED (impatiently) We don't have much time left. Why is that relevant for the liar paradox?

SOPHIA You said that the source of the paradox is self-reference. Then you also claimed that the use of Gödel's diagonal lemma shows that self-reference is involved. But it may be misleading. The Gödel and liar sentences shouldn't be compared with a person saying: 'I lost the script'; it should be compared with somebody saying: 'The last person lost the script.' If we describe the self-referentiality of the liar sentence by seeing it as a sentence that says about itself that it is not true, then we ascribe a kind of self-referentiality to it that it doesn't have.

FRED So we don't have self-reference in formal languages?

SOPHIA I didn't say that.

FRED But we get paradoxes using the weaker form of self-reference without personal pronouns.

SOPHIA Yes, but I don't fully understand what self-reference is. Not any fixed-point sentence is self-referential. In the literature people tend to describe sentences obtained via Gödel's diagonal construction as sentences saying: 'I am X.' But, as I
just said, that strikes me as unjustified.

FRED (looking at the clock on the wall) Ok, forget about the personal pronouns. Is Yablo’s paradox self-referential?

SOPHIA I don’t know.

FRED I think we have to stop here. I still think you should write an essay on the correspondence theory. So the topic for next week is: ‘Should we define truth in terms of correspondence or coherence?’

SOPHIA (gasps with her mouth and eyes wide open) Ok.

FRED (getting up and searching for a script on his desk) Bye.

SOPHIA See you next week.

SOPHIA leaves the room. Two new students come in.

FRED (nodding to the two students) David, George, your essays on the coherence and correspondence theories were brilliant.

BIBLIOGRAPHY


