This paradoxical Wittgensteinian

August 22, 2011

1 Introduction

It is mainly because of Martin that we are all very familiar with Wittgenstein's thoughts. I have to admit: sometimes I wonder whether Wittgenstein is really worth our time. At other times, however, I find it completely obvious that we should know all about him. Sometimes Wittgenstein talks nonsensical and says what he can't say, and one wonders whether he can do so. At other times he says that all truths can be expressed, though that not all that matters is expressed. Can he really claim that? Of course, some things can only be shown, but what does that mean? Wittgenstein is a paradoxical philosopher, saying a lot of nonsensical things about contradictions, but by doing so, makes us aware of new possibilities, and allowing for change.

2 Wittgenstein as paradoxical philosopher

Wittgenstein followed Russell and Moore in their rejection of the distinction that inaugurates Kants transcendental philosophy: the twofold distinction between analytic and synthetic judgements, on the one hand, and *a priori* and *a posteriori* judgements, on the other. Wittgenstein, as so many other analytic philosophers, obliterates this distinction, by assimilating the *a priori* to the analytic, and the *a posteriori* to the synthetic. The Tractatus is really an analysis of the *a priori* conditions of thought, in the Kantian tradition but it insists that the logical propositions it discusses are tautologous; and that the propositions which discuss these propositions are worse than tautologous they are nonsensical. In fact, he never gave up 'thoughts' like this. But in both his earlier and his later thought Wittgenstein's own position is highly paradoxical. This position is well summarized in his lectures on the foundations of mathematics by the maxim 'I dont have a point.' Or even by saying 'Obviously the whole point is that I must not have an opinion.'

Of course, Wittgenstein had an opinion, and he was making a point, also in his later work. For Wittgenstein is still searching for the conditions of possibility of thought and he is saying what he found out. He should have acknowledged that he is advancing opinions, and that he regards those opinions as synthetic *a priori* truths. But this is one thing, though, that Wittgenstein refuses to do.

3 The nonsense paradox

The aim of the Tractatus is to put bounds on what can be (legitimately) thought or – equivalently according to Wittgenstein – bounds on what can be (meaningfully) said.

Things that lie on the other side of the bound are simply ineffable, and to express them would be contradictory, and can thus not be done. Such 'things' can at most be *shown*.

One of those things is that facts have a certain form, the way the components are glued together. A crucial claim of Wittgenstein is that this form of a fact is not a component of this fact. Assuming that would give rise to Bradley's famous vicious regress, well known by Russell, and no doubt also by Wittgenstein. Wittgenstein claims that the form of a fact cannot be a constituent of *any* fact. But because there can be no facts with such a form as a constituent, and because we can only talk about facts, we cannot meaningfully talk about such forms. Such forms can only be shown in the use of a sentence.

Wittgenstein famously enough wanted to get rid of Russell's distinction between an object-language and its metalanguage in terms of which its semantics is expressed, how such a sentence is related to the world: Logic must take care of itself. Such a meta-language is standardly used also to express what is a meaningful sentence. But meaningful sentences must have a form, in fact, the same form as the states of affairs that they characterize. But we have just seen that we cannot meaningfully talk about forms. From this Wittgenstein consistently concludes that in our logic, or ideal language, we cannot express what are meaningful sentences.

Of course, Wittgenstein makes quite a number of remarks on what are meaningful sentences and what are not, and he also remarks that every meaningful sentence is either true or false. But on pain of contradiction, these remarks count as showing, rather than as expressing, for what is shown by them is ineffable. And indeed, this is how Wittgenstein thinks about them

6.54 Meine Sätze erläutern dadurch, dass sie der, welcher mich versteht, am Ende als uninnig erkennt, wenn der durch sie – auf ihnen – über sie hinausgestiegen ist.

(6.54) is one of the 'Sätze' of Wittgenstein's Tractatus. Thus, what Wittgenstein is 'saying' in (6.54) is not only that all he 'said' before is meaningless, but that (6.54) itself is meaningless as well. Let's forget about what he said before and concentrate only on (6.54). This can then be reduced to:

(1) This sentence is not meaningful.

What is the status of this sentence? It is either meaningless or meaningful, and in the latter case it is either true or false. Suppose it is not meaningful. In that case, what the sentence says is true. But only meaningful sentences can be true. Thus, (1) is meaningful after all. By the law of non-contradiction, we can conclude that our supposition was wrong and that (1) must be meaningful.

If (1) is meaningful, it is either true or false. Suppose it is true. Because of the meaning of (1) it follows that (1) is not meaningful. But that is in contradiction with what we assumed. Thus (1) cannot be meaningful and true.

The only alternative that is left is to assume that (1) is meaningful and false. And indeed, this doesn't give rise to contradiction. For if the sentence is false, the negation of (1) is true, meaning that (1) is meaningful.

Still, the conclusion of our reasoning is rather disturbing for Wittgenstein's project: Not only can we conclude from it that the remarks Wittgenstein makes about the nature of logic and the world are meaningful, despite his own explicit claim to the contrary, but also that what he 'shows' about it is actually false. This can hardly be what Wittgenstein intended. For this reason, I call (1) the *nonsense paradox*.

So, how, then, to understand (1)? What *is* its status? Does is really say that any talk about meaningfulness of sentences is meaningful, and sometimes false?

No, the most natural conclusion is to assume that we made an error in our 'proof' that (1) is meaningful and false. But what was the error? Perhaps some sentences are neither meaningful nor not meaningful? I can't make any sense of this proposal, however. Is it perhaps more natural to assume that some meaningful sentences are neither true nor false? No, because the whole idea of having a third option was to describe the status of meaningless sentences. But then why not say that all sentences are meaningful, but that some are neither true nor false? It doesn't help, for now (1) is still counted as false, because meaningful. In fact, it seems most natural to say that meaningless sentences are simply false, but that some meaningless sentences can still be true as well. This will be the case with paradoxical meaningless sentences.

Consider our paradoxical sentence again. Suppose it is not meaningful. In that case, what the sentence says is true. But meaningless sentences are also false. Thus, our paradoxical sentence (1) is both true and false. Of course, the negation of (1) will have the same status, and that is how it should be. Once we allow for some sentences to be both true and false, there is a danger that our language is trivial in that *every* sentence is true. Standard paraconsistent logics escape this problem by giving up on explosion: from a contradiction we cannot derive everything. But we want to be more conservative than that: explosion should stay. Instead, it is (or so we have argued in [1]) more natural to give up on transitivity of entailment. It can be shown that such a logic is quite suitable to account for paradoxes involving vagueness (Sorites) and transparent truth (the liar). But in fact, this logic works as well when we have an explicit 'meaningfulness'-predicate in our language.

4 The expressibility paradox: Lingualism

According to Wittgenstein, if something cannot be expressed, it cannot be true or false. Thus, he assumes that if something is true, it can be expressed, and thus truly expressed. Let us denote by $E\phi$ that ϕ is truly expressed (expressed and true). Wittgenstein is then committed to the following axiom: $\phi \rightarrow \Diamond E\phi$, for any ϕ. Let's call this *the thesis of expressibility*. How should the predicate, or sentential operator 'E' behave? Well, it is natural to assume that the following rules are valid: (i) $E\phi \rightarrow \phi$ and $E(\phi \wedge \psi) \rightarrow E\phi \wedge E\psi$. Unfortunately for Wittgenstein, the combination of these natural postulates is highly problematic, at least if it is assumed that one can meaningfully talk about what can be expressed, and that one can express that something can, or cannot, be expressed (which Wittgenstein denied, of course, but perhaps for no very good reasons).

On the one hand one can easily prove that from (i) and (ii) it follows that $E(\phi \wedge \neg E\phi) \rightarrow (E\phi \wedge \neg E\phi)$. By the law of non-contradiction it follows that $E(\phi \wedge \neg E\phi)$ must be false, and thus cannot be true: $\neg\Diamond E(\phi \wedge \neg E\phi)$. On the other hand, one can hardly deny that there is at least one truth that is in fact not truly expressed. Let ϕ be one of those truths. Then it is another truth that ϕ is not truly expressed: $\phi \wedge \neg E\phi$. But by Wittgenstein's expressibility thesis it follows that this latter truth can be truly expressed: $\Diamond E(\phi \wedge \neg E\phi)$. But this is in contradiction with what we have proved before. Thus we have to give up our assumption that there is a truth that is not truly expressed: $\neg(\phi \wedge \neg E\phi)$, which is equivalent to $\phi \rightarrow E\phi$. One might call this

lingualism: only those things are true that are truly expressed.

Of course, Wittgenstein is not committed to this absurd conclusion, for he doesn't allow for sentences that talk about expressibility to be true or false, and he certainly doesn't allow for embeddings of some expressibility operator. But he should have (perhaps), and the claim is only that in that case he would have been in trouble: *the expressibility paradox*.

There is another way to lingualism, which goes as follows: There are a lot of things in the universe. Many of them I have not described, or will describe in the future. Let D be a description predicate, such that Dx means that x is or will be described. From the above we can conclude that $\exists x \neg Dx$. We can talk about one of them by means of an ϵ-term: $\epsilon x \neg Dx$. Because it holds in general that $\exists x Px \rightarrow P(\epsilon x Px)$, it holds in particular that $\neg D(\epsilon x \neg D(x))$. On the other hand, we can describe something that cannot be described, for I just did: $D(\epsilon x \neg D(x))$. But now we have a contradiction: $\neg D(\epsilon x \neg D(x)) \wedge D(\epsilon x \neg D(x))$. That is, from the assumption that there are objects that are or will never be described, we can conclude a contradiction: $\exists x(Dx \wedge \neg Dx)$. From this it follows that the only objects that exists are those that are or will be described, which is a rather strong form of **lingualism**.

I am not sure how to solve the latter paradox, but I have some ideas about the former. Let us assume the following semantics for 'truly expressed' and 'possibility': $I_w(E\phi) = 1$ iff $I_w(\phi) = 1$ and there is a sentence '$@\phi$' in the language that expresses ϕ and $I_w(E\phi) = 0$ iff $I_w(\phi) = 0$ or there is no corresponding sentence, and $I_w(\Diamond\phi) = 1$ iff $\exists v \in R(w) : I_v(\phi) = 1$ and $I_w(\Diamond\phi) = 0$ iff $\forall v \in R(w) : I_v(\phi) = 0$.

We have argued before that some sentences are true and false, and we can represent this by value $\frac{1}{2}$. By the above semantics it follows that there is at least one sentence ψ such that $\Diamond E\psi$ can have at most value $\frac{1}{2}$. For the expressibility thesis to make sense for such a ψ, we want the rule $\psi \rightarrow \Diamond E\psi$ to be valid in the sense that if ψ has a value greater than or equal to $\frac{1}{2}$, $\Diamond E\psi$ will have at least value $\frac{1}{2}$. We will deny, however, that if ψ has value value 1, one can conclude that $\Diamond E\psi$ has value 1 as well.

On these assumptions it holds that for any $\phi \wedge \neg E\phi$ that is true, we can at most conclude that $\Diamond E(\phi \wedge \neg E\phi)$ has value $\frac{1}{2}$. From this we can conclude that our earlier derivation of $\neg\Diamond E(\phi \wedge \neg E\phi)$ can at most be a derivation to value $\frac{1}{2}$ (or so our expressivist is committed to). One way to account for this is to assume that necessitation is weakened in the following way: if ϕ has at least value $\frac{1}{2}$ in all models, then $\Box\phi$ has at least value $\frac{1}{2}$ in all models as well. But now we are ready, because although we have derived $\neg\Diamond E(\phi \wedge \neg E\phi)$ and $(\phi \wedge \neg E\phi) \rightarrow \Diamond E(\phi \wedge \neg E\phi)$, and that from the second it follows with contraposition that $\neg\Diamond E(\phi \wedge \neg E\phi) \rightarrow \neg(\phi \wedge \neg E\phi)$, we cannot conclude to the absurd $\neg(\phi \wedge \neg E\phi)$, i.e. $\phi \rightarrow E\phi$, because we can't chain these inferences together.

5 What an assertion shows

We have argued before that more can be expressed in our language than Wittgenstein assumed, and suggested that this can be made sense of. This doesn't mean that *everything* can be expressed. In fact, some things can't, or at least not without assuming that the 'metalanguage' is inconsistent as well. The limits of expressibility show up when we consider strengthened liar paradoxes, though in a surprising way.

We know that the liar paradox shows the limits of standard logic. Suppose that we have a third truth value as well: neither true nor false. Now the standard liar sentence (1) can be escaped:

(1) This sentence is false.

For if (1) is neither true nor false, it doesn't follow from the meaning of (1) that it is true. But now consider the so-called *extended* liar sentence (2).

(2) This sentence is not true.

If (2) is true it is not true (contradiction). If it is fase, it is true (contradiction). But if it is neither true nor false, (2) is true, and we still have a contradiction. Thus, the only escape, or so it seems, is to admit that some sentences are both *true and false*.

But if language is really very expressible, we can also say that a sentence is neither true nor false, and that it is both true and false.

(3) This sentence is neither true nor false.

(4) This sentence is both true and false.

What is the truth-value of those sentences? It is easy to see that (2) cannot have values 1 or 0, for both are explicitly denied. Indeed, (2) says of itself that it has value 'neither true nor false', and giving this value to (2) is the only available option. But if (2) is neither true nor false and the sentence says so, the sentence is true, isn't it? But that is not compatible with saying that it is neither true nor false, so this can't be an option after all.

What about (4)? Well, this time we have no problem with truth: because (4) says of itself that it is 'true and false', the sentence is true. There is nothing incompatible about this. But, unfortunately, the sentence also says of itself that it is false. But to make the sentence false, it is allowed to have at most one classical truth value, or so it seems. We seem to have similar difficulties with

(5) This sentence is true but not false. and with

(6) This is sentence is false but not true.

In fact, we seem to have difficulties with all sentences that express of themselves which truth value they have. One way out would be that our metalanguage is inconsistent as well, and that is in fact not a very strange option

Fortunately, however, we don't have to do this. Let us look at (4) again, and why it gave rise to a problem. The fallacy behind the reasoning was that we implicitly assumed that with 'false' we mean 'only false'. But, in fact, we shouldn't do that: if we say that something is false we only mean that falsity is included in the value of this something. This becomes clear if we assume that with 'it is false that ϕ' we simply mean 'It is not true that ϕ' and that negation is just standard Kleene-negation, where $\frac{1}{2}$ now models 'both true and false'. In that case, (4) (5) and (6) should be represented by

(4') $T(4') \wedge F(4')$ which means $T(4') \wedge \neg T(4')$

(5') $T(4') \wedge \neg F(4')$ which means $T(4') \wedge \neg\neg T(4')$

(6') $F(4') \wedge \neg T(4')$ which means $\neg T(4') \wedge \neg T(4')$

But one can show that in models that obey transparant truth, (4') has value $\frac{1}{2}$, (5') can have values 1 and $\frac{1}{2}$, and (6') has value $\frac{1}{2}$. Thus, there is really no problem with sentences like (4), (5) and (6).

This discussion does show something else, though. One cannot express in the language which exact truth value the sentence has: one can express that a sentence has *at least* value $\frac{1}{2}$ or *at most* value $\frac{1}{2}$, and (4') and (6') show that for at least some sentences it can be be expressed that this sentence has *exactly* value $\frac{1}{2}$, but it is impossible to express in the language that one particular sentence has *exactly* value 0, or *exactly* value 1. There is no Kripkean fixed-point model for our language with a transparent truth predicate in which this can be done.

But if we can't express that a sentence has *exactly* value 0, or *exactly* value 1, how can we know that we disagree? that I do not hold everything to be true (enough)? For my saying that ϕ is not true is consistent with the claim that it is. We need showing after all, or so it seems. But 'showing' is something different from what Wittgenstein thought: it is a pragmatic process, ruled by pragmatic principles. The one most important principle is the strongest meaning hypothesis. If we assume that within the assertion in which it is used and the further context, ϕ is consistent,[1] we interpret ϕ as being only true, i.e., as having value 1. This also holds for $\neg\phi$. Thus, the assertion of a sentence $\neg\phi$, if used in the right context, rules out the truth (even a tolerant one) of ϕ. This, then, is the way we can rule out triviality: we don't *say* that something is only true, but we rely on the hearer to conclude this from our assertion, it is *shown*.

6 Wittgenstein on contradictions

Consistency became of crucial importance mainly due to the ideas of Hilbert. Ludwig Wittgenstein was not convinced. This came out very clearly during his seminar on the Foundations of Mathematics (Diamond, 1967) in his discussions with Türing.[2]

> Wittgenstein: Think of the case of the Liar: It is very queer in a way that this should have puzzled anyone much more extraordinary than you might think... Because the thing works like this: if a man says 'I am lying' we say that it follows that he is not lying, from which it follows that he is lying and so on. Well, so what? You can go on like that until you are black in the face. Why not? It doesn't matter. ...it is just a useless language-game, and why should anyone be excited?
>
> Turing: What puzzles one is that one usually uses a contradiction as a criterion for having done something wrong. But in this case one cannot find anything done wrong.
>
> Wittgenstein: Yes and more: nothing has been done wrong, ... where will the harm come?
>
> Turing: The real harm will not come in unless there is an application, in which a bridge may fall down or something of that sort. You cannot be confident about applying your calculus until you know that there are no hidden contradictions in it.

[1] This is not without problems: we all know it is not easy to prove consistency.

[2] Diamond, C. (1967), *Wittgenstein's Lectures on the Foundations of Mathematics*, Cambridge, 1939 Cornell University Press.

Wittgenstein: There seems to me an enormous mistake there. ... Suppose I convince [someone] of the paradox of the Liar, and he says, 'I lie, therefore I do not lie, therefore I lie and I do not lie, therefore we have a contradiction, therefore 2x2 = 369.' Well, we should not call this 'multiplication,' that is all...

Turing: Although you do not know that the bridge will fall if there are no contradictions, yet it is almost certain that if there are contradictions it will go wrong somewhere.

Wittgenstein: But nothing has ever gone wrong that way yet...

Thus, according to Wittgenstein, from some systems that work pretty well, there is nothing wrong with the derivation of a contradiction. What should we do with this? Wittgenstein suggested simply not to derive anything from such a contradiction:

Wittgenstein: You might get $p \wedge \neg p$ by means of Frege's system. If you can draw any conclusion you like for it, then that, as far as I can see, is all the trouble you can get into. And I would say, "Well, then, just don't drw any conclusion from a contradiction (p. 220)

Turing realized, however, that this isn't quite enough:

Turing: But that would not be enough. For if one made that rule, one could get around it and get any conclusion which one liked without actually going through the contradiction" (p. 220)

Here is the inference Turing had in mind. Let us suppose that from a theory Γ we can derive ϕ and that we can derive $\neg\phi$: (1) $\Gamma \models \phi$, and (2) $\Gamma \models \neg\phi$. The problem is that from (1) it follows that (3) $\Gamma \models \neg\phi \rightarrow \psi$. From (2) and (3), however, we can derive ψ. Thus, from a theory that gives rise to inconsistency, everything can be inferred. We don't have to go through an explicit contradictory statement to prove this.

Fortunately, this doesn't follow in a system[3] like that of Cobreros et al:[4] though modus ponens is a valid rule, detachment is not. If one derive both ϕ and $\neg\phi$, this means that both have a value $\frac{1}{2}$ in all models of Γ. It follows that $\neg\phi \rightarrow \psi$ indeed must have a value $\geq \frac{1}{2}$, and thus follows from Γ, but this is consistent with ψ having value 0. Thus, the last inference doesn't go through.

Contradictions don't do as much harm as standard logic assumes. Sometimes contradictions are even enlightening, in particular, for the analysis of change.

But you can't allow a contraction to stand: Why not? We do sometimes use this form in our talk, of course not often – but one could imagine a technique of language in which it was a regular instrument. It might for example be said of an object in motion that it existed and did not exist in this place; change might be expressed by means of a contradiction (Wittgenstein 1967, V 8)

[3]According to Jaap van der Does, Wittgenstein would probably not have agreed with this kind of solution, referring to PU 125.

[4]Cobreros, P., P. Egre, D. Ripley, and R. van Rooij, (2011), 'Tolerant, classical, strict', *Journal of Philosophical Logic*, to appear. and Cobreros, P., P. Egre, D. Ripley, and R. van Rooij (manuscript), 'Reaching transparant truth'.