A Very Short Introduction to Formalisation and the Liar Paradox

Dr. Neil Barton
neil.barton@univie.ac.at

19 March 2019

1 The Ingredients

(1.) A Truth Predicate $Tr(x)$

For a sentence ϕ we have a way of *naming sentences* (e.g. expressions of natural language, Gödel codes) $\ulcorner\phi\urcorner$.

Definition. We then say that $Tr(x)$ is a *truth predicate for $\mathscr{L}$* iff $Tr(\ulcorner\phi\urcorner)$ is well-formed for every $\phi \in \mathscr{L}$.

(2.) Certain principles about $Tr(x)$

Let $\vdash$ be a relation of logical derivability. Then we have:

Definition. *Capture.* $\phi \vdash Tr(\ulcorner\phi\urcorner)$

Definition. *Release.* $Tr(\ulcorner\phi\urcorner) \vdash \phi$.

(3.) Self-reference

We need the ability of the language to refer to sentences of itself.

This is possible in any formal system able to encode a 'reasonable' amount of mathematics (Peano arithmetic is more than enough).[1]

Theorem. (The Diagonal Lemma) Let $\mathbf{T}$ be a theory containing "enough" arithmetic. Then for any formula $\phi(x)$ in $\mathscr{L}_\mathbf{T}$ containing just x free, there is a sentence G such that $\mathbf{T} \vdash G \leftrightarrow \phi(\ulcorner G\urcorner)$.

(4.) A reasonably strong logic

e.g. classical propositional logic. Especially important are:

Law of Excluded Middle (LEM): $\phi \vee \neg\phi$

Ex Falso Quodlibet[2] (EFQ): $\bot \vdash \phi$

Disjunction Principle (DP): If $\phi \vdash \chi$ and $\psi \vdash \chi$ then $\phi \vee \psi \vdash \chi$.

Adjunction (Adj): If $\phi \vdash \psi$ and $\phi \vdash \chi$ then $\phi \vdash \psi \wedge \chi$.

2 The Derivation

Definition. The *Liar sentence* is the following sentence in a suitable language $\mathscr{L}$:

$\lambda \dashv\vdash \neg Tr(\ulcorner\lambda\urcorner)$

We then have:

1	$Tr(\ulcorner\lambda\urcorner) \vee \neg Tr(\ulcorner\lambda\urcorner$	LEM
2	$\quad Tr(\ulcorner\lambda\urcorner)$	Case I
3	$\quad \lambda$	Release, 2
4	$\quad \neg Tr(\ulcorner\lambda\urcorner)$	Def. of λ, 3
5	$\quad \neg Tr(\ulcorner\lambda\urcorner) \wedge Tr(\ulcorner\lambda\urcorner)$	Adj., 2, 4
6	$\quad \neg Tr(\ulcorner\lambda\urcorner)$	Case II
7	$\quad \lambda$	Def. of λ, 6
8	$\quad Tr(\ulcorner\lambda\urcorner)$	Capture, 7
9	$\quad Tr(\ulcorner\lambda\urcorner) \wedge \neg Tr(\ulcorner\lambda\urcorner)$	Adj., 6, 8
10	$Tr(\ulcorner\lambda\urcorner) \wedge \neg Tr(\ulcorner\lambda\urcorner)$	DP, 1—9
11	Anything!	EFQ, 1—10

Is truth broken? Some have suggested that usual language might just be straight-up problematically inconsistent (e.g. [Eklund, 2002]). But let's think how we might do better!

3 The Fixes

Notice that in any paradox, we have three options before us:

(1.) Accept or explain away the conclusion.

(2.) Deny an assumption used in the derivation.

(3.) Deny a rule of inference used in the derivation.

3.1 Typed truth

[Tarski, 1936]'s idea; truth is to be assessed in the *metalanguage* of $\mathscr{L}$, not $\mathscr{L}$ itself.

We start with our base language $\mathscr{L} = \mathscr{L}_0$.

Define the truth predictate Tr_0 for $\mathscr{L}_0$ using the T-schema:

Definition. The *T-schema* for a language $\mathscr{L}$ is the schema of assertions:

$Tr(\ulcorner\phi\urcorner) \leftrightarrow \phi$

For every sentence ϕ of $\mathscr{L}$.

Letting $\mathscr{L}_1 = \mathscr{L}_0 \cup \{Tr_0\}$, we can then define a truth predicate for $\mathscr{L}_1$ similarly.

[1] In fact **P**rimitive **R**ecursive **A**rithmetic is enough. See [Boolos et al., 2007] for a description of $\mathbf{PRA}$.

[2] This is also sometimes known as *explosion*.

But at each stage we enforce the requirement that no language contains its own truth predicate.

Truth for $\mathscr{L}_\alpha$ can only be defined in $\mathscr{L}_{\alpha+1}$.[3]

There's no way of formulating a liar for any of these languages, the sentence $\lambda \leftrightarrow \neg Tr_\alpha(\ulcorner \lambda \urcorner)$ simply never appears: No Tr_α can apply to sentences of the same level.

Advantage. The Tarskian approach lets us carry on as normal, using classical logic.

Disadvantage. This seems far away from what we originally intended by *truth*.

3.2 Dialethism and paraconsistent solutions

A different approach: Accept that there are true contradictions! (This is known as *dialethism*.)

But what about EFQ? Surely we can't have *everything* be true.

Response: Adopt a logic that violates EFQ (so called *paraconsistent* logics).

These are often backed up by a semantics on which we have three values: True ($\top$), False ($\bot$), and Both ($\top\bot$). Priest's Logic of Paradox **LP** is one such, and violates EFQ.

Advantage. We get to keep our 'naive' conception of a single truth predicate applying to everything.

Disadvantage 1. True contradictions seem, at best, puzzling (but there is a rich literature here, e.g. [Priest et al., 2018] for a start).

Disadvantage 2. There are real problems with conditionals for many of these logics (much of the literature focusses on how to remedy this, again see [Priest et al., 2018]).

e.g. Priest's Logic of Paradox (**LP**) violates *modus ponens* since we can have $Val(\phi) = \top\bot$, $Val(\phi \to \psi) = \top\bot$, but $Val(\psi) = \bot$.

3.3 Truth-value gaps and paracomplete solutions

Instead of saying the Liar is true *and* false, maybe we should say it is *neither*.

If that's the case, then perhaps there's something wrong with the Law of Excluded Middle.[4]

[Kripke, 1975] showed how we could successively take an extension of the truth predicate, an anti-extension (for the falsity predicate), and hit a fixed-point (where truth does not change).

This motivates *Strong Kleene Logic* where we have three values: True ($\top$), False ($\bot$), and Neither (N).

We can show in this logic show that the liar sentence gets value N, and LEM fails.[5]

Advantage. We get an untyped truth predicate.

Disadvantage 1. There are real problems supplementing Strong Kleene logic with a reasonable conditional.

e.g. Capture and release both fail badly: We don't have $\neg Tr(\ulcorner \phi \urcorner) \vee \phi$ for every ϕ, and so don't have $Tr(\ulcorner \phi \urcorner) \to \phi$.

Again, there's attempted fixes for this (e.g. [Field, 2008]).

Disadvantage 2. There appear to be problems of revenge.

Of course, Kripke's theory blocks the paradoxes, but when we really think what the strengthened Liar says informally ("I'm not true!"), it seems true because the Liar isn't true, it's neither.

Letting $Fa(x)$ be the falsity predicate, if we were allowed a predicate $Ne(x)$ into the language, this thought can be formally encoded by the sentence: $\phi \leftrightarrow [Fa(\ulcorner \phi \urcorner) \vee Ne(\ulcorner \phi \urcorner)]$.

As Kripke himself put it in 1975: "The ghost of the Tarski hierarchy is still with us."

4 Conclusions

How to deal with the Liar technically is a very difficult issue!

The one thing we know: There is a "bump in the rug" to be pushed somewhere, and one has to motivate where to move it.

There's a very rich literature out there. I have left out *a lot*.

Here I have put things in *semantic* terms, and merely pointed to some *axiomatic* theories.[6]

Much of this summary has been extracted from the wonderful [Beall et al., 2017]. For the references, see the version up on the 'Blog' section of my website `https://neilbarton.net`.

[3] Eagle-eyed readers will notice that I've used a variable α here, normally associated with ordinals (which can be infinite). The reason: Tarski's construction can be extended transfinitely.

[4] This isn't necessarily so, see [Rumfitt, 2015] for a book-length treatment of the claim that non-bivalent (i.e. permitting neither true nor false sentences) semantics are consistent with the use of classical logic.

[5] In fact Strong Kleene is exactly dual to Priest's Logic.

[6] See, for example, the SEP entry [Halbach and Leigh, 2018] or the encyclopedic [Halbach, 2011].

References

[Beall et al., 2017] Beall, J., Glanzberg, M., and Ripley, D. (2017). Liar paradox. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, fall 2017 edition.

[Boolos et al., 2007] Boolos, G., Burgess, J., P., R., and Jeffrey, C. (2007). *Computability and Logic*. Cambridge University Press.

[Eklund, 2002] Eklund, M. (2002). Inconsistent languages. *Philosophy and Phenomenological Research*, 64(2):251–275.

[Field, 2008] Field, H. (2008). *Saving Truth From Paradox*. Oxford University Press.

[Halbach, 2011] Halbach, V. (2011). *Axiomatic Theories of Truth*. Cambridge University Press.

[Halbach and Leigh, 2018] Halbach, V. and Leigh, G. E. (2018). Axiomatic theories of truth. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, spring 2018 edition.

[Kripke, 1975] Kripke, S. (1975). Outline of a theory of truth. *The Journal of Philosophy*, 72(19):690–716.

[Priest et al., 2018] Priest, G., Berto, F., and Weber, Z. (2018). Dialetheism. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, fall 2018 edition.

[Rumfitt, 2015] Rumfitt, I. (2015). *The Boundary Stones of Thought: An Essay in the Philosophy of Logic*. Oxford University Press.

[Tarski, 1936] Tarski, A. (1936). The concept of truth in formalized languages. In Tarski, A., editor, *Logic, Semantics, Metamathematics*, pages 152–278. Oxford University Press.