Revision Forever!

Benedikt Löwe

Institute for Logic, Language and Computation, Universiteit van Amsterdam,
Plantage Muidergracht 24, 1018 TV Amsterdam, The Netherlands
bloewe@science.uva.nl

Abstract. Revision is a method to deal with non-monotonic processes. It has been used in theory of truth an an answer to semantic paradoxes as the liar, but the idea is universal and resurfaces in many areas of logic and applications of logic.
In this survey, we describe the general idea in the framework of pointer semantics and point out that beyond the formal semantics given by Gupta and Belnap, the process of revision itself and its behaviour may be the central features that allow us to model our intuitions about truth, and is applicable to a lot of other areas like belief, rationality, and many more.

1 Paradoxes

Paradoxes have been around since the dawn of formal and informal logic, most notably the **liar**'s paradox:

This sentence is false.

Obviously, it is impossible to assign one of the truth values *true* or *false* to the liar's sentence without a contradiction. One of the most pertinacious urban legends about the liar's paradox and related *insolubilia* is that the problem is just self-reference. But it cannot be so simple; a lot of self-referential sentences are completely unproblematic ("This sentence has five words"), and others that formally look very similar to the liar, have a very different behaviour. For example, look at the **truthteller**

This sentence is true.

As opposed to the liar, the truthteller can consistently take both the truth values *true* and *false*, but it is still intuitively problematic: there is no way we can find out whether the sentence is correctly or incorrectly asserting its own truth. The same happens with the so-called **nested liars**:

The next sentence is false,
the previous sentence is false.

Here, the assumption that the first sentence is false and the second is true is perfectly consistent, as is the assumption that the first sentence is true and the second false. If you mix the liar with a truthteller and let them refer to each other, you get the **nested mix**,

the next sentence is false,
the previous sentence is true,

which again does not allow a consistent truth value assignment.

Even though all of them are problematic, their status is subtly different and we get a rather clear picture of how and why they are different. Even more striking is the following **hemi-tautology**:

At least one of the next and this sentence is false,
both the previous and this sentence are false.

Here we get a unique consistent truth value assignment; the first sentence must be *true* and the second one *false*, and our intuition allows us to identify it accurately.[1]

In this survey, we shall discuss structural approaches based on the concept of revision due to Herzberger [He82a,He82b] and Gupta and Belnap [GuBe93] called **revision theory**. We describe revision theory both as a partial truth predicate based on revision (this is the way Gupta and Belnap phrase it in their book) and as a conceptual method. We argue that the underlying ideas of revision theory are widely applicable; the formal semantics has been reinvented independently in many areas of logic (§ 6.1), and the conceptual framework of recurrence and stability describes a wide range of phenomena (§ 6.2).

2 Pointer Semantics

In § 3, we shall describe the semantics of Herzberger, Gupta and Belnap in the simple logical language of **pointer semantics** invented by Gaifman [Ga88,Ga92]. The presentation of the system in this section is taken from [Bo03, § 5].

We shall define a propositional language with pointers $\mathcal{L}$ with countably many propositional variables p_n and the usual connectives and constants of infinitary propositional logic ($\bigwedge$, $\bigvee$, $\neg$, $\top$, $\bot$). Our language will have **expressions** and **clauses**; clauses will be formed by numbers, expressions and a pointer symbol denoted by the colon : .

We recursively define the expressions of $\mathcal{L}$:

- Every p_n is an expression.
- $\bot$ and $\top$ are expressions.
- If E is an expression, then $\neg E$ is an expression.
- If the E_i are expressions, then $\bigwedge_{i\in\mathbb{N}} E_i$ and $\bigvee_{i\in\mathbb{N}} E_i$ are expressions.
- Nothing else is an expression.

If E is an expression and n is a natural number, then $n\colon E$ is a **clause**. We intuitively interpret $n\colon E$ as "p_n states E". We can easily express all of the examples from § 1 as (sets of) clauses in this language. For instance, the liar is just the clause $0\colon \neg\mathrm{p}_0$ ("the 0th proposition states the negation of the 0th

[1] For a critical discussion of reasoning of this type, *cf.* [$\mathrm{Kr}_0$03, p. 331-332].

proposition"). The truthteller is $0\colon \mathrm{p}_0$, the nested liars are $\{0\colon \neg\mathrm{p}_1, 1\colon \neg\mathrm{p}_0\}$, the nested mix is $\{0\colon \neg\mathrm{p}_1, 1\colon \mathrm{p}_0\}$, and the hemi-tautology is $\{0\colon \neg\mathrm{p}_0 \vee \neg\mathrm{p}_1, 1\colon \neg\mathrm{p}_0 \wedge \neg\mathrm{p}_1\}$.

We now assign a semantics to our language $\mathcal{L}$. We say that an **interpretation** is a function $I\colon \mathbb{N} \to \{0,1\}$ assigning truth values to propositional letters. Obviously, an interpretation extends naturally to all expressions in $\mathcal{L}$. Now, if $n\colon E$ is a clause and I is an interpretation, we say that I **respects** $n\colon E$ if $I(n) = I(E)$. We say that I respects a set of clauses if it respects all of its elements. Finally, we call a set of clauses **paradoxical** if there is no interpretation that respects it.

Proposition 1 *The liar* $0\colon \neg\mathrm{p}_0$*, and the nested mix* $\{0\colon \neg\mathrm{p}_1, 1\colon \mathrm{p}_0\}$ *are paradoxical, the truthteller* $0\colon \mathrm{p}_0$*, the nested liars* $\{0\colon \neg\mathrm{p}_1, 1\colon \neg\mathrm{p}_0\}$ *and the hemi-tautology* $\{0\colon \neg\mathrm{p}_0 \vee \neg\mathrm{p}_1, 1\colon \neg\mathrm{p}_0 \wedge \neg\mathrm{p}_1\}$ *are non-paradoxical.*

Proof. There are four relevant interpretations here:

$$\begin{array}{ll} I_{00} & 0 \mapsto 0; 1 \mapsto 0 \\ I_{01} & 0 \mapsto 0; 1 \mapsto 1 \\ I_{10} & 0 \mapsto 1; 1 \mapsto 0 \\ I_{11} & 0 \mapsto 1; 1 \mapsto 1 \end{array}$$

It is easy to check that none of these respects the liar and the nested mix. All four interpretations respect the truthteller, and the interpretations I_{01} and I_{10} respect the nested liars. In the case of the hemi-tautology, the only respecting interpretation is I_{10}. q.e.d.

So, if the truthteller and the nested liars are non-paradoxical, does that mean that they are not problematic? Well, both I_{01} and I_{10} are interpretations of the nested liars, but the interpretations disagree about the truth values of both p_0 and p_1 and therefore do not allow any determination of truth. The situation is quite different for the hemi-tautology where there is exactly one respecting interpretation. We call a set of clauses Σ **determined** if there is a unique interpretation respecting Σ. With this notation, the truthteller and the nested liars are non-paradoxical but also non-determined, and the hemi-tautology is determined.

In [Bo02, §§ 5&6], Bolander investigates self-referentiality and paradoxicality in order to highlight that these two notions are related but there can be self-reference without paradox and paradox without self-reference. The framework of pointer semantics described so far is perfectly fit to making these claims precise.

Let Σ be a set of clauses. Then we can define the **dependency graph** of Σ by letting $\{n\,;\,\mathrm{p}_n \text{ occurs in some clause in } \Sigma\}$ be the set of vertices and defining edges by

$$nEm \text{ if and only if } \mathrm{p}_m \text{ occurs in } X \text{ for some } n\colon X \in \Sigma.$$

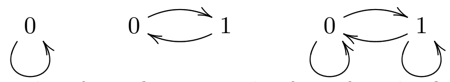

Fig. 1. Dependency graphs of our five examples from § 1: the first graph is the dependency graph for the liar and the truthteller, the second is the one for the two nested examples, and the third is the one for the hemi-tautology.

With this definition, we get the following dependency graphs for our five examples as depicted in Figure 1.

We now call a set of clauses Σ **self-referential** if there is a loop in the dependency graph of Σ. With this definition, it is obvious that self-reference does not imply paradoxicality; the clause $0\colon \mathrm{p}_0 \vee \neg\mathrm{p}_0$ shares the simple loop as a dependency graph with the liar and the truthteller, but the interpretation $I(0) = 1$ respects it. Yablo [Ya93] gave an example for the converse of this simple fact:

Proposition 2 (Yablo) *Let* $E_n := \bigwedge_{i>n} \neg\mathrm{p}_i$ *and* $\Upsilon := \{n\colon E_n \ ; \ n \in \mathbb{N}\}$. *Then* Υ *is not self-referential, but paradoxical.*

Proof. The dependency graph of Υ is $\langle \mathbb{N}, < \rangle$, so it doesn't contain any loops.

Let I be an interpretation respecting Υ. If for any $n \in \mathbb{N}$, we have $I(n) = 1$, then $1 = I(n) = I(\bigwedge_{i>n} \neg\mathrm{p}_i)$, so we must have that $I(i) = 0$ for all $i > n$. That means that $0 = I(n+1) = I(\bigwedge_{i>n+1} \neg\mathrm{p}_i)$, whence there must be some $i^* > n+1$ such that $I(i^*) = 1$. But this is a contradiction.

So, $I(n) = 0$ for all n. But then $I(E_0) = I(\bigwedge_{n>0} \neg\mathrm{p}_n) = 1 \neq 0 = I(0)$. Contradiction. q.e.d.

3 Revision

So far, our analysis did not involve revision at all – everything was solely based on the static picture given by the set of clauses. Revision theory now adds a rather natural idea of revision along the pointers established by the clauses. From now on, we shall assume that all sets of clauses Σ satisfy a simple **consistency condition**: If $n\colon E \in \Sigma$ and $n\colon F \in \Sigma$, then $E = F$. If Σ is a set of clauses, then we can define the **revision operator** on interpretations I by

$$\delta_\Sigma(I)(n) := I(E)$$

where E is the unique expression such that $n\colon E \in \Sigma$. This can now be used to recursively define a **revision sequence** of interpretations from an initial interpretation I (called "hypothesis" in revision theory) as

$$\begin{aligned} I^{\Sigma,0} &:= I \\ I^{\Sigma,n+1} &:= \delta_\Sigma(I^{\Sigma,n}). \end{aligned}$$

We call an interpretation J **Σ-recurring** if there is some I such that there are infinitely many n with $J = I^{\Sigma,n}$ and we call it **Σ-stable** if there is some I and some n such that for all $k > n$, we have $J = I^{\Sigma,k}$.

Proposition 3 *Let Σ be a set of clauses and I an interpretation. Then I respects Σ if and only if I is Σ-stable.*

Proof. Obviously, "I respects Σ" is equivalent to $\delta_\Sigma(I) = I$. q.e.d.

Let us check our examples from § 1. For the liar and the truthteller, relevant interpretations are just one bit ($I(0) = 0$ and $I(0) = 1$). For the liar, both interpretations are recurring, but none of them is stable. For the truthteller, both are recurring and stable. For the two nested examples, we have four relevant interpretations whose revision sequences are as follows:

nested mix:	0: $\neg p_1$	0 1 1 0 0 ···	0 0 1 1 0 ···
	1: p_0	0 0 1 1 0 ···	1 0 0 1 1 ···
	0: $\neg p_1$	1 1 0 0 1 ···	1 0 0 1 1 ···
	1: p_0	0 1 1 0 0 ···	1 1 0 0 1 ···
nested liars:	0: $\neg p_1$	0 1 0 1 0 ···	0 0 0 0 0 ···
	1: $\neg p_0$	0 1 0 1 0 ···	1 1 1 1 1 ···
	0: $\neg p_1$	1 1 1 1 1 ···	1 0 1 0 1 ···
	1: $\neg p_0$	0 0 0 0 0 ···	1 0 1 0 1 ···

For the nested mix, all four interpretations are recurring, but none of them is stable; for the nested liars, all of them are recurring, but only `01` and `10` are stable.

Analysing the revision sequences for the hemi-tautology gives us a unique stable interpretation `10` and two more recurring interpretations `00` and `11` as described in Figure 2.

hemi-tautology:	0: $\neg p_0 \vee \neg p_1$	0 1 0 1 ···	0 1 1 1 ···
	1: $\neg p_0 \wedge \neg p_1$	0 1 0 1 ···	1 0 0 0 ···
	0: $\neg p_0 \vee \neg p_1$	1 1 1 1 ···	1 0 1 0 ···
	1: $\neg p_0 \wedge \neg p_1$	0 0 0 0 ···	1 0 1 0 ···

Fig. 2. The revision sequences for the hemi-tautology.

All of this conforms with the analysis of § 2, but doesn't add any new insights. However, the revision approach can add new insights in the case that there is no unique stable solution. For this, let us consider the following example that we shall call **nested liars with two observers**:

The second sentence is false,
the first sentence is false,
exactly one of the first two sentences is true,
exactly one of the first three sentences is true.

Intuition tells us that exactly one of the first two sentences should be true, and therefore the third sentence should be true and the fourth sentence should be false. (Again, we point the reader to Kremer's debate [$\mathrm{Kr}_0$03, p. 331-332] concerning the dangers of applying ordinary reasoning to sets of sentences with self-reference.) The natural language sentences can be translated into a set of clauses as follows:

$$
\begin{aligned}
&0\colon \neg \mathrm{p}_1\\
&1\colon \neg \mathrm{p}_0\\
&2\colon (\mathrm{p}_0 \vee \mathrm{p}_1) \wedge (\neg \mathrm{p}_0 \vee \neg \mathrm{p}_1)\\
&3\colon \bigvee_{i \in 3} \mathrm{p}_i \;\wedge\; \neg \bigvee_{\substack{i \neq j\\ i,j \in 3}} (\mathrm{p}_i \wedge \mathrm{p}_j)
\end{aligned}
$$

They give rise to the revision sequences depicted in Figure 3, establishing `0110` and `1010` as the two stable interpretations, and `1100` and `0000` as recurring, yet unstable.

0	0 1 0 1 ···	0 1 0 1 ···	0 1 0 1 ···	0 1 0 1 ···
1	0 1 0 1 ···	0 1 0 1 ···	0 1 0 1 ···	0 1 0 1 ···
2	0 0 0 0 ···	0 0 0 0 ···	1 0 0 0 ···	1 0 0 0 ···
3	0 0 0 0 ···	1 0 0 0 ···	0 1 0 0 ···	1 1 0 0 ···
0	0 0 0 0 ···	0 0 0 0 ···	0 0 0 0 ···	0 0 0 0 ···
1	1 1 1 1 ···	1 1 1 1 ···	1 1 1 1 ···	1 1 1 1 ···
2	0 1 1 1 ···	0 1 1 1 ···	1 1 1 1 ···	1 1 1 1 ···
3	0 1 0 0 ···	1 1 0 0 ···	0 0 0 0 ···	1 0 0 0 ···
0	1 1 1 1 ···	1 1 1 1 ···	1 1 1 1 ···	1 1 1 1 ···
1	0 0 0 0 ···	0 0 0 0 ···	0 0 0 0 ···	0 0 0 0 ···
2	0 1 1 1 ···	0 1 1 1 ···	1 1 1 1 ···	1 1 1 1 ···
3	0 1 0 0 ···	1 1 0 0 ···	0 0 0 0 ···	1 0 0 0 ···
0	1 0 1 0 ···	1 0 1 0 ···	1 0 1 0 ···	1 0 1 0 ···
1	1 0 1 0 ···	1 0 1 0 ···	1 0 1 0 ···	1 0 1 0 ···
2	0 0 0 0 ···	0 0 0 0 ···	1 0 0 0 ···	1 0 0 0 ···
3	0 0 0 0 ···	1 0 0 0 ···	0 0 0 0 ···	1 0 0 0 ···

Fig. 3. The revision patters of nested liars with two observers.

While the four recurring interpretations disagree about the truth values of p_0, p_1, and p_2, all of them agree that p_3 should receive value 0. Therefore, even in the absence of a unique solution, we can get information out of the revision procedure and define a partial truth predicate.

If Σ is a set of clauses and $n\colon X \in \Sigma$, then we say that p_n is **stably true** (**recurringly true**) if for every stable (recurring) interpretation I, we have $I(n) = 1$. Similarly, we define notions of being **stably false** and **recurringly false**. The difference between the stable partial truth predicate and the recurring partial truth predicate is roughly the difference between the Gupta-Belnap systems $\mathbf{S}_0$ and $\mathbf{S}_n$.[2] Gupta and Belnap argue [GuBe93, Example 5A.17] that $\mathbf{S}_0$ is not good enough to capture intuitions. The systems $\mathbf{S}^*$ and $\mathbf{S}^\#$ proposed by Gupta and Belnap [GuBe93, p. 182 & 191] are refinements of these systems. The differences hardly matter for simple examples of the type that we are covering in this paper.

Proposition 4 *In the nested liars with two observers, the fourth sentence is recurringly false.*

Proposition 4 sounds like a success for the revision theoretic analysis of the concept of truth, as it gives a prediction or analysis for a truth value that coincides with the intuition. However, it is important to note that our reasoning used to intuitively determine the truth value of the fourth sentence used the fact that the third sentence seemed to be intuitively true. But the revision analysis is less informative about the third sentence: it is neither recurringly true nor recurringly false, but stably true. This phenomenon (with a different example) was the topic of the discussion between Cook and Kremer in the journal *Analysis* [Co02,Kr$_0$03,Co03] and will be discussed in detail in § 4.

4 Fully revised sequences and the Cook-Kremer debate

In a dispute in the journal *Analysis* [Co02,Kr$_0$03,Co03], Roy Cook and Michael Kremer debated whether the revision-theoretic analysis of self-referential sentences yields intuitive or counterintuitive readings. Both Cook and Kremer focussed on what we called "recurring truth" in the last section.

The hemi-tautology from § 1 is a special case of the following set of clauses. As usual, denote by $\binom{k}{n}$ the set of k-element subsets of $n = \{0, ..., n-1\}$. For every positive natural number n, the set Σ_n has the n clauses

$$k\colon \bigvee_{X \in \binom{k+1}{n}} \bigwedge_{i \in X} \neg \mathrm{p}_i$$

(for $k < n$), *i.e.*, "there are at least $k+1$ many false sentences". If n is odd, Σ_n is paradoxical, if n is even, then it has a unique respecting interpretation, *viz.* the one in which sentences $0, ..., \frac{n}{2}$ are true and the rest false. The original example in [Co02] is Σ_4, the hemi-tautology is the example used in [Kr$_0$03] and is Σ_2 in the above notation. Analysing the revision sequences in Figure 2, we get:

Proposition 5 *In the hemi-tautology, neither of the sentences receives a recurring truth value.*

Proof. The recurring interpretations are `10`, `00` and `11`, and so they agree on neither of the truth values. q.e.d.

[2] *Cf.* [GuBe93, p. 123 & 147].

Cook [Co02] contrasts the partial truth predicate of recurring truth as calculated Proposition 4 with our intuitive expectations of a favoured interpretation `10` for the hemi-tautology, and considers this a failure of the revision theoretic analysis.

It is surprising that neither Cook nor Kremer mention that this phenomenon has been observed by Gupta and Belnap. They discuss this in a slightly less transparent example [GuBe93, Example 6C.10]:

The third sentence is true,
It is true that the third sentence is false,
One of the first two sentences is false,

formalized as

$$\{0\colon \mathrm{p}_3, 1\colon \neg\mathrm{p}_3, 2\colon \mathrm{p}_1, 3\colon \neg\mathrm{p}_0 \vee \neg\mathrm{p}_2\},$$

where intuition suggests that `1001` should be the only solution. Analysing the revision sequences, we find that `1001` is the only stable interpretation, but `0101`, `1011`, and `1000` are recurring, and thus none of the four truth values is determined in the Gupta-Belnap revision semantics defined *via* recurring interpretations.

Gupta and Belnap deal with this situation with their notion of "fully varied" revision sequences. We extend the sequences from sequences indexed with natural numbers to transfinite sequences indexed with ordinal numbers.[3] Given a limit ordinal λ, we say that a revision sequence $s = \langle I_\xi\,;\,\xi < \lambda\rangle$ **coheres** with an interpretation I if the following two conditions are met:

1. If for some $\xi < \lambda$ and all $\eta > \xi$, we have $s_\eta(n) = 1$, then $I(n) = 1$.
2. If for some $\xi < \lambda$ and all $\eta > \xi$, we have $s_\eta(n) = 0$, then $I(n) = 0$.

So, going back to the case of $\lambda = \omega$, if the value of n has stabilized after a finite number of revisions, then an interpretation must agree with this value in order to cohere. For those n that flip back and forth infinitely many times, the value of $I(n)$ can be both 0 or 1. Looking at the hemi-tautology as an example, we get four revision sequences as in Figure 2:

$$\begin{matrix} 0\,1\,0\,1\cdots \\ 0\,1\,0\,1\cdots \end{matrix}$$

$$\begin{matrix} 0\,1\,1\,1\cdots \\ 1\,0\,0\,0\cdots \end{matrix}$$

$$\begin{matrix} 1\,1\,1\,1\cdots \\ 0\,0\,0\,0\cdots \end{matrix}$$

$$\begin{matrix} 1\,0\,1\,0\cdots \\ 1\,0\,1\,0\cdots \end{matrix}$$

[3] The "forever" in the title of this paper is an allusion to this extension of the process of revision into the transfinite.

The ones starting with `01` and `10` stabilize on `10`, and so only `10` is a coherent interpretation for them. The other two flip back and forth infinitely many times in both slots, and so every interpretation is coherent with those.

Using the notion of coherence, we can now define the notion of a transfinite revision sequence. If Σ is a set of clauses and δ_Σ is the revision operator derived from Σ in the sense of § 3, then a sequence $s = \langle I_\xi \,;\, \xi < \lambda \rangle$ of interpretations is called a **transfinite revision sequence** if $I_{\xi+1} = \delta_\Sigma(I_\xi)$ and I_ϱ coheres with $s{\restriction}\varrho$ for limit ordinals ϱ. Note that for a fixed interpretation I_0 there can be different transfinite revision sequences starting with I_0.

Gupta and Belnap call a transfinite revision sequence **fully varied** if every interpretation coherent with it occurs in it [GuBe93, p. 168]. For the hemi-tautology, the sequences starting with `01` and `10` are fully varied; the only coherent interpretation is `10` and it occurs in them. The other two sequences are not fully varied, as `01` and `10` cohere with them, but don't occur. However, we can transfinitely extend them to the four revision sequences

$$\begin{array}{ll}
0\,1\,0\,1\cdots & 0\,1\,1\,1\cdots \\
0\,1\,0\,1\cdots & 1\,0\,0\,0\cdots \\[4pt]
0\,1\,0\,1\cdots & 1\,1\,1\,1\cdots \\
0\,1\,0\,1\cdots & 0\,0\,0\,0\cdots \\[4pt]
1\,0\,1\,0\cdots & 0\,1\,1\,1\cdots \\
1\,0\,1\,0\cdots & 1\,0\,0\,0\cdots \\[4pt]
1\,0\,1\,0\cdots & 1\,1\,1\,1\cdots \\
1\,0\,1\,0\cdots & 0\,0\,0\,0\cdots\,,
\end{array}$$

characterized by their values at 0 and the ordinal ω as `00/01`, `00/10`, `11/01`, and `11/10`. All of these sequences (of length $\omega \cdot 2$) are fully varied, and together with the sequences starting with `01` and `10`, they are essentially the only fully varied sequences.

We can now define a new notion of recurrence. Given a transfinite revision sequence s of length λ for a set of clauses Σ, we say that I is recurring in s if for all $\xi < \lambda$ there is some $\eta > \xi$ such that $s_\eta = I$. Based on this notion, we say that p_n is **transfinitely true** (**transfinitely false**) if for all fully varied transfinite revision sequences s and all interpretations I that are recurring in s, we have $I(n) = 1$ ($I(n) = 0$).

Proposition 6 *The first sentence of the hemi-tautology is transfinitely true, the second is transfinitely false.*

This alternative analysis arrives at the intuitive expectations by enforcing additional constraints on the notion of a revision sequence. Cook implicitly acknowledges this possible defense of the revision analysis when he says

> "The Revision Theorist might ... formulat[e] more complex revision rules than the straightforward one considered here, ones that judged the sentences [of the hemi-tautology] as non-pathological. [Co03, p. 257]"

The fact that there are so many different systems of revision theory, all with slightly different requirements on the sequences or variations of the semantic predicate, each of them with some other set of advantages and disadvantages, is raising a concern: we are trying to model a phenomenon as central as truth; if revision theory is a fundamental tool to understanding it, shouldn't it provide answers that do not depend on such minor details?

One possible way out of trouble would be to get rid of the idea that a theory of truth needs to define a partial truth predicate. Revision theory gives a rich analysis of what happens, yielding patterns of behaviour of truth values. Instead of superposing these patterns into a single (partial) interpretation as is done by the notions of "stable truth", "recurring truth" and "transfinite truth", we could understand the revision analysis as the description of what is going on:

The liar is problematic as there are no stable interpretations, the truthteller is because there are two conflicting ones. This difference explains how they are different types of problems for the theorist of truth – collapsing it into a uniform partial truth function (which would give the value "undefined" to both the liar and the truthteller) clouds a rather clear conceptual picture. We propose to think of the sequences and their behaviour as the real analysis of truth without the definition of a partial truth predicate; the fact that `10` is the only stable interpretation for the hemi-tautology is good enough to explain our intuitions with the set of sentences.[4]

It is this approach to revision sequences that we believe to be a powerful tool for explaining intuitions with truth, much more than the different axiomatic systems proposed by various authors in order to deal with inadequacies of earlier definitions. We shall continue this discussion in § 6.2.

5 An aside: "And what is the connection to Belief Revision?"

In the community of applied and philosophical logic, the word "revision" is much closer associated to the area of *belief revision* and *belief update* than to the revision theory described in § 3. In 2002, I gave a talk on the complexity of revision-theoretic definability at the annual meeting of the Pacific Division of the *American Philosophical Association* with the title "Where does the complexity of revision come from?",[5] and received questions from philosophical logicians asking about the complexity of *belief revision* in the style of [Li97,Li00].

Is the use of the phrase "revision" in both areas just an equivocation? Do the two underlying concepts of revision ("update of belief states in light of changing reality" and "update of truth value in a formal system") have nothing to do with each other?

[4] Note that by Proposition 3, this is equivalent to saying that `10` is the only interpretation that respects the hemi-tautology, so here the pointer semantics approach and the revision approach are just two different ways of looking at the same phenomenon.

[5] The results of this talk have in the meantime been published as [KüLöMöWe05].

In this section, we shall give a rough sketch of why revising belief states may be incorporated into the framework described in §§ 2 and 3. Since this is a side issue here, we cannot do justice to these questions here.

In belief revision and update, we have an ordinary propositional language and consider sets of formulae as **belief sets**. Based on new information about the true state of the world, we may get inconsistent intermediate stages of belief sets which we then have to update in order to reach a consistent belief set again. This is the main paradigm of an enormous amount of literature in philosophy, logic and artificial intelligence.[6]

The most basic example is the following: an agent believes that p and $p \rightarrow q$ are true, but then learns that $\neg q$ is true. The belief set has to be updated to either $\{p, \neg q, \neg(p \rightarrow q)\}$ or $\{\neg p, \neg q, p \rightarrow q\}$. Of course, which one is the correct update will depend on the context.

We believe that revision theory as described in § 3 can provide a partial semantics for belief update procedures in general, but will only develop this idea for the simple examples given above here. Given a belief set Λ and some new fact represented by a propositional variable, we can assign a set of clauses in our language $\mathcal{L}$ as follows:

Let Λ^* be the set of propositional variables occurring in a formula in Λ and let $\pi : \Lambda^* \rightarrow \mathbb{N}$ be an injective function with coinfinite range. We can think of π as associating an $\mathcal{L}$-variable p_n to each element of Λ^*. Clearly, π naturally extends to all elements of Λ.

In a second step, we take an injective function $\pi^* : \Lambda \rightarrow \mathbb{N}$ such that $\mathrm{ran}(\pi) \cap \mathrm{ran}(\pi^*) = \varnothing$. If $n \in \mathrm{ran}(\pi) \cup \mathrm{ran}(\pi^*)$, we define a clause $n\colon E$ where

$$E := \begin{cases} \mathrm{p}_n, & \text{if } n \in \mathrm{ran}(\pi), \\ \pi(\varphi), & \text{if } \varphi \in \Lambda \text{ and } \pi^*(\varphi) = n. \end{cases}$$

This defines the set Σ of $\mathcal{L}$-clauses associated to Λ.

In our given example, this would be

$$\{0\colon \mathrm{p}_0, 1\colon \mathrm{p}_1, 2\colon \mathrm{p}_0 \rightarrow \mathrm{p}_1\}.$$

The dependency graph of our set of clauses is

$$\begin{tikzcd} & 2 \arrow[dl, bend right] \arrow[dr, bend left] & \\ 0 & & 1. \end{tikzcd}$$

The key difference between the setting of revision theory and that of belief update is that the new fact that triggers the update is given a special status: if the initial belief set is $\{p, p \rightarrow q\}$ and we learn $\neg q$ as a fact, then we don't want to disbelieve this fact in order to remedy the situation.

We fix some $n \in \mathrm{ran}(\pi)$ and some truth value $b \in \{0, 1\}$ for this n, assuming that the new fact that we learned corresponds to p_n or $\neg \mathrm{p}_n$. An $\langle n, b\rangle$**-interpretation** is a function $I : \mathbb{N} \rightarrow \{0, 1\}$ that satisfies $I(n) = b$.

[6] As a token reference, we mention [Gä92], in particular the introduction.

0: p_0	$0\,0\,0\cdots$	$0\,0\,0\cdots$
1: p_1	$0\,0\,0\cdots$	$0\,0\,0\cdots$
2: $\mathrm{p}_0 \to \mathrm{p}_1$	$0\,1\,1\cdots$	$1\,1\,1\cdots$
0: p_0	$1\,1\,1\cdots$	$1\,1\,1\cdots$
1: p_1	$0\,0\,0\cdots$	$0\,0\,0\cdots$
2: $\mathrm{p}_0 \to \mathrm{p}_1$	$0\,0\,0\cdots$	$1\,0\,0\cdots$

We see that `001` and `100` are the only stable interpretations. Taking our remarks at the end of § 4 seriously, we shall not use this to define a partial truth function (which would say that p_1 is recurringly false and the others have no assigned truth value), but instead look at the two stable interpretations and see that

$$\{\mathrm{p}_0, \neg\mathrm{p}_1, \neg(\mathrm{p}_0 \to \mathrm{p}_1)\} \text{ and } \{\neg\mathrm{p}_0, \neg\mathrm{p}_1, \mathrm{p}_0 \to \mathrm{p}_1\}$$

are the two possible outcomes for the belief set after belief update.

6 The ubiquity of revision

In the abstract, we mentioned that revision is a concept that is "universal and resurfaces in many areas of logic and applications of logic". It comes in two very different flavours as discussed at the end of § 4: as formal Gupta-Belnap semantics defining a partial truth predicate on the basis of revision sequences, and in the wider sense as a conceptual framework for analysing our intuitions about truth and circularity. So far, we have argued that revision plays a rôle in the analysis of paradoxes and *insolubilia*, and that the approach may be applied to belief revision. In this section, we shall lay out how the general ideas can be extended to yield applications in other areas. We split the discussion into applications of the Gupta-Belnap semantics and applications of the wider scope.

6.1 Independent developments of Gupta-Belnap semantics

The crucial mathematical element to the Gupta-Belnap truth predicate as defined in § 3 (as "recurring truth") is the following: we have a set of nonmonotonic processes assigning a function $I : \mathbb{N} \to \{0, 1\}$ to each ordinal. While monotonic processes give rise to fixed points and thus allow us to talk about an "eventual value", nonmonotonicity forces us to be inventive here. The processes give rise to a notion of *recurrence*, and we can define

$$T_{\mathrm{GB}}(n) := \begin{cases} 0 \text{ if for all recurrent } I\text{, we have } I(n) = 0, \\ 1 \text{ if for all recurrent } I\text{, we have } I(n) = 1, \\ \uparrow \text{ otherwise.} \end{cases}$$

This is a general idea to integrate the process of revision into a single definition, and Gupta and Belnap are not the only ones who came up with this idea. Essentially the same semantics was developed independently by Stephan Kreutzer in [Kr$_1$02] for his partial fixed point logics on infinite structures. Also Field's *revenge-immune solution* to the paradoxes from [Fi03] is based on ideas very similar to the Gupta-Belnap semantics.[7]

Widening the scope to other types of transfinite processes, cognate ideas can be found in the limit behaviour of infinite time Turing machines as defined by Hamkins and Kidder [HaLe00][8] and definitions of game labellings for nonmonotone procedures for game analyses in [Lö03].

6.2 The wider scope

Despite the fact that the general ideas have found applications in many plases, there are several problems with Gupta-Belnap semantics as a theory of truth. As mentioned, there are many variants of formal systems with different properties, thus raising the question of how to choose between them. The Cook-Kremer debate discussed in § 4 is an indication for the problems generated by this. The revision-theoretic definitions are also relatively complicated, leading (in the language of arithmetic) to complete Π^1_2 sets, in the case of using fully revised sequences even Π^1_3 sets [We03a, Theorem 3.4]. This is too complicated for comfort, as is argued in [We01, p. 351] and [LöWe01, § 6].

As we have discussed in § 4, the conceptual idea of analysing the nonmonotonic process by looking at the behaviour of interpretations under revision rises above all this criticism. The problems associated with the arbitrariness and complexity of the Gupta-Belnap are related to the fact that the full analysis has to be condensed into one partial truth predicate. Allowing both `01` and `10` as stable solutions of the nested liars is much more informative than integrating these two solutions into undefined values.

This attitude towards allowing several possibilities as analyses should remind the reader of game-theoretic solution concepts. In game theory, Nash equilibria are not always unique. This connection between revision semantics and game theory has been observed by Chapuis who gives a sketch of a general theory of rationality in games based on revision analyses in his [Ch03]. We see Chapuis' work as an interesting approach compatible with the spirit of the analysis of belief update discussed in § 5, and would like to see more similar approaches to revision in various fields of formal modelling.

References

[Bo02] Thomas **Bolander**, Self-Reference and Logic, **Phi News** 1 (2002), p. 9-44

[Bo03] Thomas **Bolander**, Logical Theories for Agent Introspection, PhD thesis, Technical University of Denmark 2003

[7] Welch has proved in [We03b] that the set of *ultimately true* sentences in the sense of [Fi03] coincides with the set of stable truths in the sense of Herzberger.

[8] This similarity was pointed out by the present author in [Lö01] and used by Welch in [We03a] to solve the limit rule problem of revision theory. *Cf.* also [LöWe01].

[Ch03] André **Chapuis**, An application of circular definitions: Rational Decision, *in:* Benedikt Löwe, Wolfgang Malzkorn, Thoralf Räsch (*eds.*), Foundations of the Formal Sciences II: Applications of Mathematical Logic in Philosophy and Linguistics, Rheinische Friedrich-Wilhelms-Universität Bonn, November 10-13, 2000, Dordrecht 2003 [Trends in Logic 17], p. 47-54

[Co02] Roy T. **Cook**, Counterintuitive consequences of the revision theory of truth, **Analysis** 62 (2002), p. 16-22

[Co03] Roy T. **Cook**, Still counterintuitive: a reply to M. Kremer, "Intuitive consequences of the revision theory of truth", **Analysis** 63 (2003), p. 257-261

[Fi03] Hartry **Field**, A revenge-immune solution to the semantic paradoxes, **Journal of Philosophical Logic** 32 (2003), p. 139-177

[Ga88] Haim **Gaifman**, Operational Pointer Semantics: Solution to Self-referential Puzzles I, *in:* Moshe Vardi (*ed.*), Proceedings of the 2nd Conference on Theoretical Aspects of Reasoning about Knowledge, Pacific Grove, CA, March 1988, Morgan Kaufmann, San Francisco 1988, p. 43–59

[Ga92] Haim **Gaifman**, Pointers to Truth, **Journal of Philosophy** 89 (1992), p. 223–261

[Gä92] Peter **Gärdenfors** (*ed.*), Belief revision, Cambridge University Press 1992 [Cambridge Tracts in Theoretical Computer Science 29]

[GuBe93] Anil **Gupta**, Nuel **Belnap**, The Revision Theory of Truth, Cambridge MA 1993

[HaLe00] Joel David **Hamkins**, Andy **Lewis**, Infinite time Turing machines, **Journal of Symbolic Logic** 65 (2000), p. 567-604

[He82a] Hans G. **Herzberger**, Naive Semantics and the Liar Paradox, **Journal of Philosophy** 79 (1982), p. 479–497

[He82b] Hans G. **Herzberger**, Notes on Naive Semantics, **Journal of Philosophical Logic** 11 (1982), p. 61–102

[$\mathrm{Kr}_0$03] Michael **Kremer**, Intuitive consequences of the revision theory of truth, **Analysis** 62 (2002), p. 330-336

[$\mathrm{Kr}_1$02] Stephan **Kreutzer**, Partial Fixed-Point Logic on Infinite Structures, *in:* Julian C. Bradfield (*ed.*), Computer Science Logic, 16th International Workshop, CSL 2002, 11th Annual Conference of the EACSL, Edinburgh, Scotland, UK, September 22-25, 2002, Proceedings, Berlin 2002 [Lecture Notes in Computer Science 2471], p. 337-351

[KüLöMöWe05] Kai-Uwe **Kühnberger**, Benedikt **Löwe**, Michael **Möllerfeld**, Philip **Welch**, Comparing inductive and circular definitions: parameters, complexities and games, **Studia Logica** 81 (2005), p. 79-98

[Li97] Paolo **Liberatore**, The complexity of iterated belief revision, *in:* Foto Afrati, Phokion Kolaitis (*eds.*), Database theory—ICDT '97, Proceedings of the 6th International Conference held in Delphi, January 8-10, 1997, Springer-Verlag 1997 [Lecture Notes in Computer Science 1186], p. 276-290

[Li00] Paolo **Liberatore**, The complexity of belief update, **Artificial Intelligence** 119 (2000), p. 141-190

[Lö01] Benedikt **Löwe**, Revision sequences and computers with an infinite amount of time, **Journal of Logic and Computation** 11 (2001), p. 25-40; *also in:* Heinrich Wansing (*ed.*), Essays on Non-Classical Logic, Singapore 2001 [Advances in Logic 1], p. 37-59

[Lö03] Benedikt **Löwe**, Determinacy for infinite games with more than two players with preferences, **ILLC Publication Series** PP-2003-19

[LöWe01] Benedikt **Löwe**, Philip D. **Welch**, Set-Theoretic Absoluteness and the Revision Theory of Truth, **Studia Logica** 68 (2001), p. 21-41

[We01] Philip D. **Welch**, On Gupta-Belnap Revision Theories of Truth, Kripkean fixed points, and the Next Stable Set, **Bulletin of Symbolic Logic** 7 (2001), p. 345-360

[We03a] Philip D. **Welch**, On Revision Operators, **Journal of Symbolic Logic** 68 (2003), p. 689–711

[We03b] Philip D. **Welch**, Ultimate Truth *vis à vis* stable truth, *preprint*, November 7, 2003

[Ya93] Stephen **Yablo**, Paradox without self-reference, **Analysis** 53 (1993), p. 251-252