Psychophysiological approach to the Liar paradox : Jean Buridan's virtual entailment principle put to the test

Synthese manuscript No.
(will be inserted by the editor)

Psychophysiological approach to the Liar paradox. Jean Buridan's virtual entailment principle put to the test.

Konrad Rudnicki · Piotr Łukowski

Received: date / Accepted: date

Abstract This article presents an empirical examination of the consequences of the *virtual entailment principle* proposed by Jean Buridan to resolve the Liar paradox. This principle states that every sentence in natural language implicitly asserts its own truth. Adopting this principle means that the Liar sentence is not paradoxical but false, because its content is contradictory to what is virtually implied. As a result, humans should perceive the Liar sentence the same way as any other false sentence. This solution to the Liar paradox received criticism for making *ad hoc* claims about the natural language. However, thanks to modern advancements in psychophysiology, it became possible to empirically investigate if the human brain really perceives the Liar sentence like a false sentence. We designed and conducted an experiment to examine brain activity in response to true sentences, false sentences and self-referential sentences (including the Liar and the Truthteller). Our results provide support for the Buridan's hypothesis and show that the Liar sentence is processed by the human brain identically to false sentences, while the Truthteller sentence is perceived identically to true sentences. This agrees with predictions derived from the *virtual entailment principle* and supports the idea that humans think with the logic of truth – a logic for which the truth is a designated value of its adequate semantics[1].

Konrad Rudnicki
Department of Communication Science, University of Antwerp.
E-mail: konrad.rudnicki@uantwerpen.be

Piotr Łukowski
Institute of Philosophy, Department of Logic and Methodology of Science, University of Łódź. E-mail: lukowski@uni.lodz.pl

[1] This research is supported by the National Science Centre of Poland - grant Nr 2015/17/B/HS1/02332 for the project: "Epistemological aspects of using the content implication connective as a tool of formalizing natural language expressions."

Keywords Liar paradox · Jean Buridan · entailment · relativism · ERP · N400 · experimental philosophy · neurophilosophy

1 Introduction

Paradoxes are the unwanted children of philosophy. Whenever a paradox is discovered, philosophers immediately start working to recognize its cause and make it disappear. For "The Liar paradox", that work started 2500 years ago and continues to this day. As a result, there are multiple competing theories attempting to solve it. Some of these theories address the definition of truth, while others change the boundaries of what is allowed in natural language, or even replace the classical formal logic with another type of logic altogether (see Martin 1984 or Łukowski 2011 for an overview). Most of these solutions are purely formal and cannot be easily studied in empirical terms. However, there is one class of solutions to the Liar paradox that can be empirically investigated. The approaches belonging to this class assume that the Liar sentence is contradictory and therefore false. Advocates for that type of solutions can be found in the Middle Ages (e.g. Jean Buridan, Thomas Bradwardine, Albert of Saxony) (Rahman et al. 2008) as well as in the modern times (e.g. Charles Sanders Peirce, Arthur Prior, Eugene Mills, Neil Lefebvre and Melissa Schelein, Piotr Łukowski) (Rahman et al. 2008, Prior 1961, Lefevbre and Schelein 2005, Łukowski 2011). Most of these modern solutions were derived from the so-called 'Buridan's thesis' (i.e. *the virtual entailment principle*).

The virtual entailment principle devised by Jean Buridan in the XIV-th century proposes that every sentence of the natural language implicitly asserts its own truth (Buridan trans. 2001). Accepting this principle causes the Liar sentence to become contradictory and therefore false, which means that it is no longer a paradox. As a result, Buridan explains, we should treat the Liar sentence as if it was simply a falsehood. The lack of empirical evidence supporting this claim caused it to be criticized for its *ad hoc* nature (Benetreau-Dupin 2014). Fortunately, modern psychophysiology delivers tools that make it possible to examine if there is merit in accepting the *virtual entailment principle*. We designed and performed an experiment to determine if the Liar sentence is really perceived by the human brain like a false sentence. In this article we report the result of that experiment and discuss its implications for the contemporary philosophical considerations on truth.

We divide this article into 7 sections. In §2 we start by introducing the liar paradox itself and describe in detail how Jean Buridan comes to his solution (i.e.*the virtual entailment principle*). In §3 we present how the logic with content implication esteems the Buridan's principle. Sections §4 and §5 explain how contemporary research in psycholinguistics already provided some preliminary evidence for the Buridan's thesis and propose a new experiment to test it directly. In §6 we demonstrate the results of that experiment and finally, in

§7 we discuss how our results fit within the existing theories of truth and what are their implications.

2 Buridan's solution of the Liar paradox

This statement is false. This paradoxical sentence is an example of an alethic, self-falsifying statement, most often represented as:

$$(L) : L\ is\ false.$$

Attempts to determine the truth value of L lead to a loop of reasoning. If we assume that L is true, then it is true what is said by L, and so L is false. If L is false, then, what is said by L is the case, therefore L is true, and the loop starts over. The existence of such sentence raises questions about our theory of truth and the utility of logic in the natural language and of natural language in logic. If a theory of truth employs the law of non-contradiction and the principle of bivalence, then it seems necessary to somehow determine the truth value of the Liar sentence under that theory. Furthermore, if natural language as a formal system generates antinomies it might not be well-suited for accurately describing the reality. Because of these problems, philosophers and logicians have already worked for over two millennia trying to solve the Liar paradox. In this project we will focus on the solution devised by Jean Buridan, who proposed a *virtual entailment principle* to remove the Liar paradox from the natural language.

In order to determine what is the truth value of the Liar sentence one needs a general criterion for truth that can be applied to any sentence. Jean Buridan's solution to the Liar paradox rests on the *conception of truth by supposition* which states that a sentence is true when subject and predicate supposit (i.e. stand for) the same (Benetreau-Dupin 2014). Traditionally, an issue with this conception arises when considering the truth value of the Liar sentence. Assuming "L" to be true leads to the conclusion that "L" and "false" supposit for the same thing, therefore "L" is false. At the same time, assuming "L" to be false leads to the conclusion that "L" and "false" do not supposit for the same thing, therefore "L" is true. Buridan was aware that his *conception of truth by supposition* was not sufficient to handle the Liar sentence, so he introduced his *virtual entailment principle* (Klima 2018). He wrote: *"(. . .) every proposition virtually implies another proposition in which the predicate 'true' is affirmed of the subject which supposits for [the original proposition]; and I say 'virtually implies' in the way in which the antecedent implies that which follows from it."* (Buridan trans. 2001). How does that resolve the paradox? If every sentence implicitly asserts its own truth, then the sentence L_1:

$$(L_1) : L_1\ is\ false.$$

is equivalent to the sentence L_2:

$$(L_2) : L_2\ is\ true.\ and\ L_2\ is\ false$$

Sentence L_2 is not paradoxical, but false which means that by extension one can conclude that L_1 is also false. As a result, the Liar paradox does not exist anymore in natural language. This solution to the paradox was popularized and formally described by Arthur Prior (1961) (cf. Uckelman 2012).

Critical voices raised concerns about the *virtual entailment principle* and pointed out that it is *ad hoc* to add such a broad assumption regarding the whole natural language only to deal with self-referential sentences (Read 2002, 2006). Supporters of Buridan argue that even though *virtual entailment principle* indeed deals with paradoxes of self-reference it is a natural consequence of the *conception of truth by supposition* and does not claim anything arbitrary (Benetreau-Dupin 2014, Hughes 1982, Klima 2018). In our study we will contribute to the debate whether the *virtual entailment principle* is warranted by addressing the problem from empirical perspective. In particular, we will investigate the consequences of the *virtual entailment principle* and test the hypothesis that the Liar sentence is perceived by humans like a false sentence.

3 The logic with content implication and Buridan's principle

Let $\mathcal{L}_{CCL} = (L_{CCL}, \neg, \wedge, \vee, \rightarrow, \leftrightarrow, :)$ be an extension of the classical propositional logic by the new connective ":". The *Contentual Classical Logic* (CCL) defined on $\mathcal{L}_{CCL}$ is an axiomatic extension of the classical propositional logic where the added axioms are in the following form (Łukowski 2011):

$A_{1:}$ $((\alpha : \beta) \wedge (\beta : \delta)) \rightarrow (\alpha : \delta))$
$A_{2:}$ $(\alpha \wedge \beta) : \alpha$
$A_{3:}$ $(\alpha \wedge \beta) : (\beta \wedge \alpha)$
$A_{4:}$ $\alpha : (\alpha \wedge \alpha)$
$A_{5:}$ $((\alpha : \beta) \wedge (\beta : \alpha)) \rightarrow ((\neg\alpha : \neg\beta) \wedge (\neg\beta : \neg\alpha))$
$A_{6:}$ $((\alpha : \beta) \wedge (\beta : \alpha) \wedge (\delta : \gamma) \wedge (\gamma : \delta)) \rightarrow (((\alpha \# \delta) : (\beta \# \gamma)) \wedge ((\beta \# \gamma) : (\alpha \# \delta)))$,
$for\ \# \in \{\rightarrow, \leftrightarrow, :\}$
$A_{7:}$ $((\alpha : \beta) \wedge (\delta : \gamma)) \rightarrow ((\alpha \# \delta) : (\beta \# \gamma))$, $for\ \# \in \{\wedge, \vee\}$
$A_{8:}$ $(\alpha : \beta) \rightarrow (\alpha \rightarrow \beta)$

Modus Ponens (MP) $\{\alpha \rightarrow \beta, \alpha\} \vdash \beta$ remains the only inference rule. A semantic adequate for CCL is the class of all so-called CCL-models, i.e., matrices $\mathcal{M} = (\mathcal{A}, D)$, such that $\mathcal{A} = (A, -, \cap, \cup, \Rightarrow, \Leftrightarrow, \supset)$ is an algebra similar to

$\mathcal{L}_{CCL}, D$ is a nonempty subset of A for all $a, b \in A$,

1. $a = a \cap a$
2. $a \cap b = b \cap a$
3. $a \cap (b \cap c) = (a \cap b) \cap c$
4. $-a \in D \; iff \; a \notin D$
5. $a \cap b \in D \; iff \; a \in D \; and \; b \in D$
6. $a \cup b \in D \; iff \; a \in D \; or \; b \in D$
7. $a \Rightarrow b \in D \; iff \; a \notin D \; or \; b \in D$
8. $a \supset b \in D \; iff \; a = b \cap c, \; for \; some \; c \in A$

Semantic inference is defined in a standard way: $X \vDash_{CCL} \alpha \; iff$ for any CCL-model $\mathcal{M} = (\mathcal{A}, D) \; and \; v \in Hom(\mathcal{L}_{CCL}, \mathcal{A}) \; v(\alpha) \in D, \; if \; for \; any \; \beta \in X, v(\beta) \in D$.

"*p says what is said by q*" or shortly "*p says q*" is a intended reading of the sentence $p : q$. A semantic interpretation of the sentence $p : q$ shows that the sense of q is a part of the sense of p. It coincides with the fact that, the only CCL-tautologies with the connective of the content implication as a main functor are the following formulas:

$$\alpha : \alpha \; and \; (\alpha_1 \wedge ... \wedge \alpha_n) : \alpha_i \; for \; i \in \{1, ...n\}$$

It means that the sense of a sentence is constructed with senses of other sentences. Moreover, the holistic paradigm is not excluded here.

It seems worth of noticing that CCL is a non-Fregean logic in the Suszko's sense. At the beginning of the 1970th, Roman Suszko together with his coworker Stephen Bloom constructed the *Sentential Calculus with Identity* (SCI) (Bloom and Suszko 1975). This calculus invalidated the Frege's axiom assuming that all true sentences have one and the same reference, the truth, and all false sentences have one and the same reference, the falsehood. This has been achieved by the only one non-truth-functional connective of the sentential identity, defined by the following class of axioms:

$A_{1\equiv} \; \alpha \equiv \alpha$
$A_{2\equiv} \; (\alpha \equiv \beta) \rightarrow (\neg\alpha \equiv \neg\beta)$
$A_{3\equiv} \; ((\alpha \equiv \beta) \wedge (\gamma \equiv \delta)) \rightarrow ((\alpha \S \gamma) \equiv (\beta \S \delta)), \; for \; \S \in \{\wedge, \vee, \rightarrow, \leftrightarrow, \equiv\}$
$A_{4\equiv} \; (\alpha \equiv \beta) \rightarrow (\alpha \rightarrow \beta)$

Indeed, an adequate SCI semantic requires a class of models with more than two semantic correlates. In accordance with the semantic interpretation intended by Suszko semantic correlates of the SCI-model should be understood as situations – two sentences are identical, if they have one and the same semantic correlate, i.e. both sentences expressed the same situation. Today, an increasingly popular interpretation is to understand semantic correlates as the contents of sentences. There is a close relation between CCL and SCI. The

formula: $((\alpha : \beta) \wedge (\beta : \alpha)) \leftrightarrow (\alpha \equiv_c \beta)$ defines a sentential identity $\equiv$ which is the Suszko's connective enriched by three axioms:

$$(\alpha \wedge \alpha) \equiv_c \alpha$$

$$(\alpha \wedge \beta) \equiv_c (\beta \wedge \alpha)$$

$$(\alpha \wedge \beta) \wedge \gamma \equiv_c \alpha \wedge (\beta \wedge \gamma)$$

From the point of view of the subject of this paper, the most important CCL-theses is, just mentioned, $\alpha : \alpha$ easily inferred by A_1:, A_2: and A_4:. Despite its trivial form, this formula is not trivial because it expresses the famous Buridan's principle. It prevents us from forgetting that we "breathe" with the logic of truth and not falsehood. As it was shown by Jan Woleński, the Liar sentence is antinomial only on the logic of truth, since on the logic of falsehood it does not lead to contradiction, unlike the Truthteller sentence which is antinomial only on the logic of falsehood, while on the logic of truth it does not lead to contradiction (Woleński 1993). Let us recall that a given logic is the logic of truth, if the truth is a designated value of its adequate semantics. For example, in the case of two intuitionistic logics, the Heyting logic is a logic of truth, while the Brouwerian logic, a logic of falsehood – in the case of the Brouwerian logic, a standard sense of all connectives (especially conjunction, disjunction, and co-implication) is kept only, if the falsehood is a designated value of models for this logic. As Woleński showed, the sense of the Liar sentence and the Truthteller sentence depends on the logical value in which we are thinking about these sentences. Thinking in truth is a reflex, nobody gives it a second thought. However, it is crucial for the Liar/Truthteller problem. Thus, the Liar sentence has its well-known sense only on the logic of truth. In other words, the Liar sentence says that it is false, only if we are thinking of it in the logic of truth. In the logic of falsehood, the sense of the Liar sentence would be opposite. CCL is the logic of truth because $\alpha : \alpha$ is a CCL-tautology. In the light of the desired meaning of the sentence with ":" as a main functor, a sense of the Liar sentence L is expressed by the sentence

$$L : \neg L$$

Since every sentence says what is said by this sentence, also $L : L$. Thus, in accordance with the Buridan's wish by A_1, A_2, A_4

$$L : (L \wedge \neg L)$$

In the light of the semantic interpretation, a sense *v(L)* of L is the following

$$v(L) = v(L \wedge \neg L) \cap c, \ \textit{for some } c$$

It means that the content of the permanently false sentence $L \wedge \neg L$ is a part of the sense of L. More specifically, let $\mathcal{M} = (\mathcal{A}, D)$ be a CCL-model, $v \in (\mathcal{L}_{\mathrm{CCL}}, \mathcal{A})$ be a homomorphism such, that $v(L) = a_0 \in A$. Since L is the Liar sentence, a formula $L : \neg L$ is satisfied in $\mathcal{M}$ by v. It means that,

$a_0 \supset -a_0 \in D$. By the eighth condition of a CCL-model, $a_0 = -a_0 \cap c$, for some $c \in A$. There are two cases, either: $a_0 \in D$ or $a_0 \notin D$.

If $a_0 \in D$, then also $-a_0 \cap c \in D$. By the fifth condition, $-a_0 \in D$, and so by the fourth one, $a_0 \notin D$ - a contradiction. Thus, the sentence L cannot be satisfied in $\mathcal{M}$ by v.

Let $a_0 \notin D$. By the fourth condition, $-a_0 \in D$ and so $a_0 = -a_0 \cap c \notin D$, for some $c \in A$. Thus, $-a_0 \notin D$ or $c \notin D$. Since $-a_0 \in D$, so $c \notin D$. Thus, a reasoning will be successfully completed by finding such a false sentence z, that $L : z$. But, an existence of such a sentence was already proved, it is $L \wedge \neg L$. Thus, let $c = v(z) = v(L \wedge \neg L) \notin D$ - no contradiction.

Summarizing, L is a false sentence not being true. Although L says that L is false, L is not true because L says that is false only as a true sentence. It means that L says much more that it is false, it says that it is true and false at the same time. It formally confirms the Buridan's informal statement.

Prior's solution of the Liar antinomy belongs to the class of those formal approaches which assumed that every uttered and well understood sentences are treated as true. Arthur Prior developed his idea of the calculus with proposition-forming functors of propositions (Prior 1961).

4 Truth and falsehood in psycholinguistics

Psycholinguistic research already provides some clues indicating that Buridan might have been right.

First, in a number of experiments, Gilbert et al. (1990, 1993) presented participants with a series of sentences, which were either true or false. True sentences were displayed in one color, false sentences in another. One group of participants read the sentences without any interference. The other had to simultaneously count numbers backwards in memory (which interferes with sentence comprehension). Then, after a delay, participants viewed some of these sentences again, without any colors, and answered a question: *Was this sentence true or false?* Participants who were not interfered had no problems with remembering which sentences were true, and which were false. At the same time those who had to count in memory sometimes remembered wrong and answered incorrectly. However, they only made one type of mistake. They recalled false sentences as true, but almost never made the mistake of recalling true sentences as false. This suggests, that when in a situation of doubt and incomplete information, the brain has the tendency to treat all sentences as true. Gilbert et al. (1990, 1993) intended his experiments to resolve the debate initiated by Spinoza (trans. 1982) and Descartes (trans. 1984) if the truth evaluation happens before or after sentence comprehension. Gilbert et al. (1990, 1993) interpreted his results in favor of Spinoza and concluded that truth evaluation precedes comprehension. Further research clarified that this is not exactly the case. Psychophysiological studies showed that the processes of comprehension precede truth evaluation, but also interact with them (Hagoort et al. 2004, Wiswede et al. 2013). However, the experiments by Gilbert

et al. (1990, 1993) showed that our brain tends to treat every information as true when it has no reasons to treat it otherwise. This result agreed with the *virtual entailment principle*.

The fact that humans perceive new information as true if they do not possess any previous contradictory data is reflected in the way our brain processes that information. It was shown that comprehending obviously false sentences requires cognitive effort and higher brain activity (Hagoort et al. 2004, Marques et al. 2009, Wiswede et al. 2013). This is expected, because for false sentences additional information needs to be retrieved from memory and compared against the content of those sentences (Marques et al. 2009). Areas of the brain active during false sentences comprehension are known to be active in demanding reasoning tasks and processing self-generated information, which supports the idea that identifying falsehood involves finding a contradiction between the sentence information and information stored in memory (Marques et al., 2009). What is remarkable is that this higher effort for false sentences is present both at the stage of comprehension and at the stage of truth evaluation (Hagoort et al. 2004, Wiswede et al. 2013). How could this be? How can falsehood cause effort at the stage when it was not yet established by the brain to be false?

The technique used to extract brain activity during sentence comprehension is called electroencephalography (EEG). More specifically, event-related potential (ERP) technique is used to study the average activity of large groups of neurons following stimuli presentation. Electrodes placed on the participant's scalp record the mean bio-electrical activity generated by the cortex of the brain. The result is a time locked average signal that human brain produces in response to a certain class of stimuli, e.g. true and false sentences. Past research have shown that there are stable temporal patterns within this signal, and differences in these patterns correspond to certain cognitive phenomena. For example, Hagoort et al. (2004) studied differences in the brain activity evoked by true sentences, sentences false because of a semantic error and sentences false because of a factual error. Their results initiated a new stream of research concerned with differences in how the brain processes true and false sentences.

When a human brain perceives language, it constantly tries to predict what will be heard or read next and anticipates the upcoming words. For example, if we read:

The sun is a ...

the brain already anticipates the word "*star*" and prepares that word in a neural semantic network, in order to conserve energy. If a different, surprising word appears at the end of that sentence, it requires more effort from the brain to comprehend it. Then, more brain activity can be seen at the stage of "*semantic integration*" (i.e. comprehending the meaning of the whole sentence) (Hagoort et al. 2004). This stage happens approximately 400 ms from the onset of the last word in a sentence. As a result, psychophysiological studies were able to identify how differently the brain signal looks for obviously false and

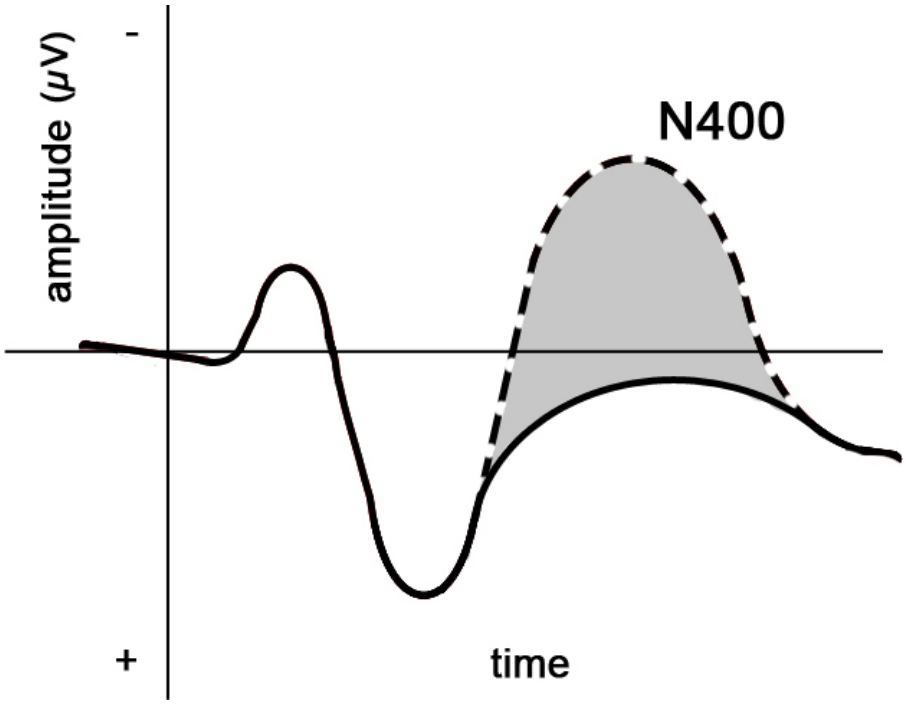

Fig. 1 Schematic visualization of brain activity after reading natural language sentences. Vertical line signifies the onset of the last word in a sentence. Solid black line represents typical signal for true sentences. Dashed line represents how signal looks for obviously false sentences. This negative deflection in brain activity is called N400 and signifies an increased cognitive effort in integrating the meaning of a sentence.

true sentences. False sentences cause a deeper negative signal 400 ms from the last word in a sentence. This effect is called N400, where "N" stands for *negative*. We present the visualization of this effect in Figure 1. The phenomenon where some words violate the expectations of the reader and cause the N400 is also called "*semantic mismatch*" or "*semantic incongruence*" (Dudschig et al. 2016). In other words, an unexpected word causes cognitive effort in integrating the meaning of a sentence. Then, more effort for false sentences is also seen at the stage of truth evaluation, approximately 1000ms from the onset of last word in a sentence (Wiswede et al. 2013). What does it mean?

It means that cognitive effort at the stage of comprehension (N400) prompts our brain to pay special caution and spend more resources to verify the content of the sentence which caused the effort. As a result, false sentences cause effort at multiple stages. First, when they hit the brain with an unexpected word, and later when the brain tries to resolve why did that word appear there. If *virtual entailment principle* is true and every sentence implicitly asserts its own truth which makes Liar sentence false, then human brain should also react to the Liar sentence like it reacts to false sentences.

It is crucial that we focus on the stage of sentence comprehension and not sentence verification. Effort at the level of comprehension is caused by failed unconscious anticipations of the brain, while effort at the level of verification it is caused by an ongoing conscious reasoning process. It is likely that the Liar sentence would cause effort at the level of reasoning - which is what everyone experiences while trying to consciously determine if the Liar is lying or telling the truth. However, that type of effort would not indicate that the Liar sentence is perceived as false, because conscious cognitive effort can be evoked by many different things, not only falsity (i.e. unexpectedness) of a statement.

If the *virtual entailment principle* is true then the word "*false*" at the end of the Liar sentence should be unexpected for the brain and constitute a "*semantic mismatch*". This should cause deeper N400, just like false sentences do. In contrast, the Truthteller sentence ("*This sentence is true.*") should not be a mismatch, because its content agrees with the *virtual entailment principle* and can be anticipated by the brain.

5 Experimental design and procedure

The hypothesis of our study can be formulated in terms of a theory of truth:

H: The liar sentence is perceived like a false sentence.

This general hypothesis needs to be made more specific by being operationalized (i.e. formulated in a way that can be directly, empirically studied). In our study the specific form of this hypothesis is therefore formulated as:

H: Reading the liar sentence causes the same N400 amplitude as reading a false sentence.

Finally, the Liar sentence has another property that needs to be accounted for in the experimental design – it is a self-referential sentence. In psychology, the term "self-referential" denotes sentences that address the person who reads them. For example, "*I am tall.*" However, in philosophy, saying that the Liar sentence is self-referential means that its content concerns the sentence itself. Self-referentiality in this sense is a separate phenomenon from paradoxicality. A sentence can be self-referential and paradoxical like the Liar sentence, but also self-referential and true (e.g. "*This sentence is in English*"), as well as self-referential and false (e.g. "*This sentence is in Dutch.*") To make sure that any differences in brain activity observed during processing of the Liar sentence are due to its truth value and not due to its self-referentiality we designed an experiment including other self-referential sentences along with normal sentences.

5.1 Participants

30 students from the SWPS University of Social Sciences and Humanities in Warsaw participated in the study. All participants had normal or corrected to

Table 1 Types of sentences used in the experiment

Sentence Type	Example
Liar sentences	*This statement is false*
Truthteller sentences	*This statement is true*
Self-referential false sentences	*This sentence is in German*
Self-referential sentences with semantic error	*This phrase is wet*
Self-referential true sentences	*This statement is in English*
Normal false sentences	*Sun is a planet*
Normal sentences with semantic error	*Sun is a crime*
Normal true sentences	*Sun is a star*

normal vision and declared no history of neurological or psychiatric disorders or medication. Participants were required to refrain from taking psychoactive substances on the day of the experiment. Every participant provided written informed consent according to the Declaration of Helsinki twice. First, before the procedure, where the study was introduced as: "*Cognitive aspects of the natural language – an ERP study*", then again after the procedure where the study title was corrected to: "*Cognitive aspects of the Liar paradox – an ERP study.*" Study title on the first consent form was altered to make sure that the participants do not pay special attention to paradoxical sentences during stimuli presentation. All participants were compensated with 30PLN after the experiment. This study was approved by the Committee for Ethics in Scientific Research of the SWPS University.

5.2 Procedure

8 classes of sentences in Polish language were created (see Table 1). Each class consisted of 35 unique sentences, except for the truthteller group which consisted of 22. This difference was caused by a significant lack of synonyms for the word: "*truth*" in Polish, and because in evoked-potential experiments every presented sentence should be unique. As a result, 267 sentences were presented, forming 267 trials.

Participants were instructed to silently read the sentences, which were presented word-by-word on a computer screen. The exact procedure of sentences presentation is showed in Figure 2. Every word was presented for a total of 150ms + 25ms for every letter in that word. This was done to accommodate the time needed to read words of different lengths (Nieuwland and Kuperberg 2008). Between every word a 300ms blank screen was presented. The last word in every sentence was presented for a fixed time of 300ms, because the studied evoked potentials were averaged from the onset of the last word. Every sentence was followed by a blank screen for 1000ms, then fixation cross for 1000ms and again a blank screen for 1000ms. All sentences were presented in random order in blocks of 10.

After every 10 sentences participants were allowed to take a break from reading and resume when they wish by pressing the space button. During the whole procedure bioelectrical signals generated by the participants' brains'

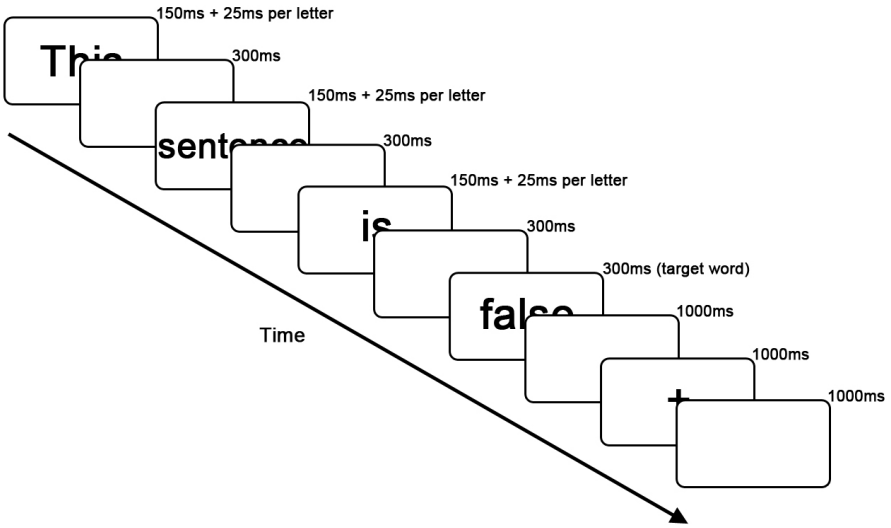

Fig. 2 Trial structure.

in response to sentence presentation were recorded with an electroencephalograph. The parameters of the signal recording are reported in the appendix. After the stimuli presentation participants answered three control questions:

1. "This sentence is false." In your opinion this sentence is: a) False b) True c) Neither

2. Did you know what is the Liar paradox before this experiment? (YES/NO)

3. Have you ever studied the problem of the Liar paradox? (e.g. lectures, educational videos) (YES/NO)

6 Results

Detailed results of the statistical analysis are reported in the appendix.

To statistically assess whether the Liar sentences cause the same brain activity as false sentences we have used the Neyman-Pearson paradigm of hypotheses testing (Neyman and Pearson 1933). It is the most widespread paradigm, used in all exact sciences for falsifying hypotheses. In this paradigm two contradictory hypotheses are constructed: a null hypothesis (H_0) and an alternative hypothesis (H_1). The aim of the statistical test is to reject (i.e. falsify) one of these hypotheses. Based on the Neyman-Pearson lemma a statistical test is constructed to calculate the probability of erroneously rejecting the null hypothesis (H_0). If this probability is lower than the arbitrarily chosen significance level α, then a decision is made to reject the null hypothesis (H_0) and keep the alternative hypothesis (H_1). Most often used significance level is 5%, which means that scientists keep the alternative hypothesis (H_1) when the probability of that being a mistake is lower than 5% (Stigler 2008). In our study there are two sets of detailed hypotheses to be tested.

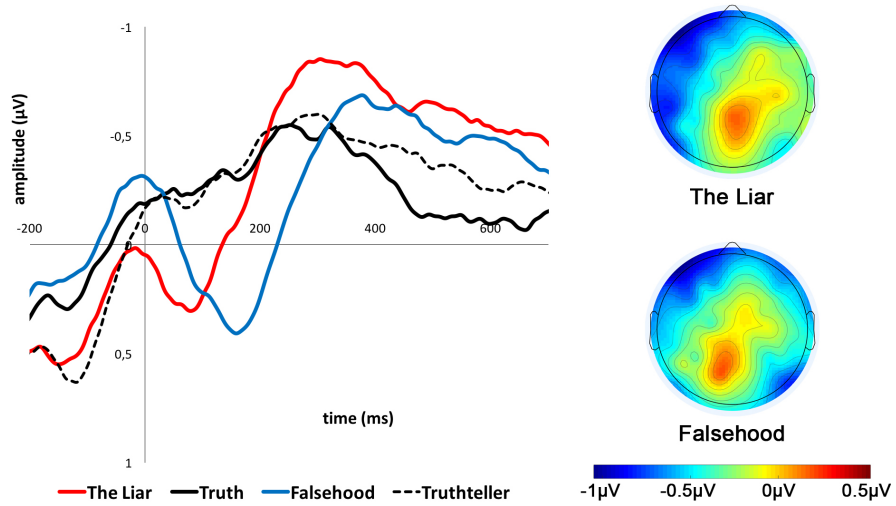

Fig. 3 Event-related potentials on representative electrodes for self-referential sentences. Spline-interpolated isovoltage maps of the scalp display the topographic distributions of the N400 component during 350-400ms interval. Signal was exponentially smoothed for visualization. Smoothing constant = 0.98

Before we determined if Liar sentences had been processed like false sentences, we needed to make sure that in our experiment false sentences were processed differently from true sentences. Past findings already confirmed this to be true, but it was crucial for us to replicate this effect in our study. Therefore, the first set of hypotheses is:

H_0:Reading false sentences causes the same N400 amplitude as reading true sentences.

H_1: Reading false sentences causes higher N400 amplitude than reading true sentences.

The second set of hypotheses to be tested is directly related to the main goal of our study:

H_0':Reading the Liar sentences causes the same N400 amplitude as reading true sentences.

H_1': Reading the Liar sentences causes higher N400 amplitude than reading true sentences.

If the statistical tests will allow us to reject both null hypotheses (H_0 and H_0'), then we can say that the Liar sentence is processed like a false sentence.

The results of our experiment allow us to confirm our research hypothesis and keep both alternative hypotheses. We present the obtained results in Figures 3 and 4.

We have replicated the past results and found that false sentences cause more effort than true sentences at the stage of comprehension. The probability that we make a mistake by saying that is 1.3%, as calculated from our sample.

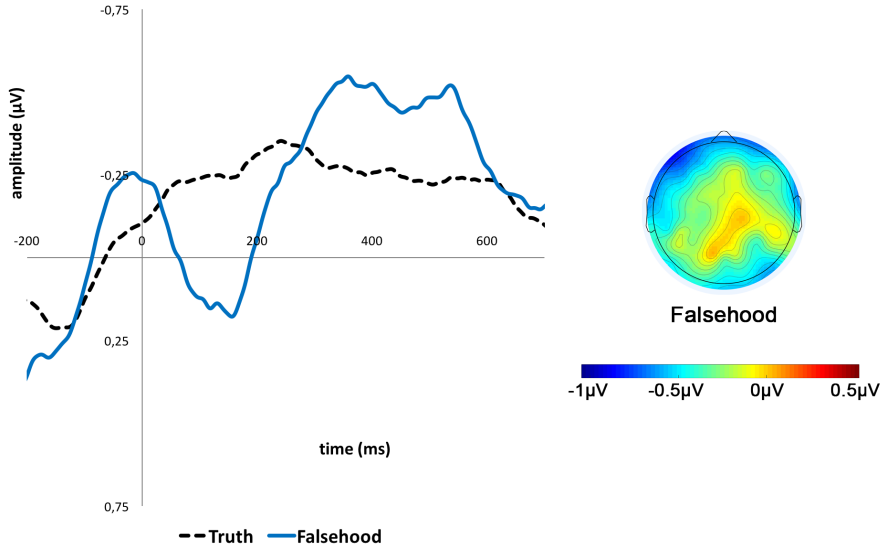

Fig. 4 Event-related potentials on representative electrodes for normal sentences. Spline-interpolated isovoltage maps of the scalp display the topographic distribution of the N400 component during 350-400ms interval. Signal was exponentially smoothed for visualization. Smoothing constant = 0.98

With regard to the main aim of our study, we have found that Liar sentences cause significantly more effort than true sentences. The probability that we make a mistake by saying that is 0.8%, as calculated from our sample.

Furthermore, we have checked if Liar sentences are somehow different from false sentences. We have found that the signal produced in response to Liar sentences is statistically identical to that caused by false sentences.

Finally, to definitely confirm that it is the paradox that causes the observed effect, and not the property of a sentence that *"it says something about its own truth value"*, we calculated if the Truthteller sentences differed from true sentences. They did not. Truthteller sentences evoked the same signal as true sentences.

7 Discussion

The aim of this study was to empirically investigate the consequences of the *virtual entailment principle* for the Liar sentence, as proposed by Jean Buridan in his conception of truth by supposition (Buridan trans. 2001). By studying brain activity we demonstrated that the Liar sentence is perceived by the human brain in the same fashion as false sentences. This result provides evidence supporting the *virtual entailment principle* and suggests that the Liar paradox

may be resolved with it. The fact that humans perceive Liar sentence like a falsehood has some implications for logic and truth theory.

Because we studied the human brain activity, our results reflect the reality of a human language use regardless of the adopted theory of truth. **We do not demonstrate that the Liar sentence *is false*. We demonstrate that the human brain reacts to the Liar sentence like it does to false sentences.** Jean Buridan wrote that his principle means that every sentence "*virtually implies*" its own truth (Buridan trans. 2001). It is debatable if this means implicature which always requires an agent making the interpretation or perhaps *entailment* which does not require an agent to perceive the sentence (Sauerland and Stateva 2007). Here we work in a paradigm where the language-user plays the main role. Furthermore, our study operates on two important hidden assumptions:

1. Human brain operates according to principles, and some of these principles can be uncovered and formally described as a "theory of truth" embedded in the human brain architecture.

2. A useful theory of truth should have predictive power with regard to the real world.

The more a given theory of truth agrees with those assumptions (or similar ones), the more our results suggest that such a theory should include the *virtual entailment principle*. That places our experiment within the relativistic view on truth, where adjectives pertaining to the truth-value are *assessment-sensitive* and an agent is required to perceive them to assess them (MacFarlane 2014).

The fact that our study needs to accept relativism is also its main theoretical limitation. To say that someone "perceives the Liar sentence to be false" is to say something about his mental states (i.e. qualia). Qualia cannot be subjected to an empirical study. All that science is able to study are correlates (i.e. epiphenomena) of the qualia. In our experiment we study a very specific physiological correlate of the fact that someone perceives a sentence to be false. When humans hear a sentence, they continuously try to predict every upcoming word they might hear next. This "predictive coding" helps to perceive language faster and conserves energy. Whenever humans hear a sentence in which that predictive coding fails, their brain spends more energy to understand its meaning. This failure to meet predictions of the brain can have various reasons, not only falsity. If someone believes in something false, their brain will react with effort upon hearing the truth. In principle, we study the mismatch between someone's expectations of what they will hear and what they have actually heard (the N400). It is our interpretation that the reason why the Liar sentence causes N400 is that it is understood to be false. However, there are some arguments in support of that interpretation.

First, for self-referential sentences the expectations are set almost entirely by the context when they are being read. What can we expect at the end of a sentence: "This sentence is…?" The reader of this article may expect for instance: *"in English"*, *"in an article"* or *"in black font."* In our experiment we confirm that self-referential sentences which abide by these expectations

do not cause effort in the brain (N400), while self-referential sentences which fail them, do. A question arises: *"What would be the expectation of the brain with regard to the truth value of a self-referential sentence?"* Jean Buridan attempted to answer this question and proposed that we should expect every sentence to implicitly assert its own truth. There are two alternatives to this hypothesis. Either, sentences implicitly assert their own falsity, which is an absurd hypothesis and can be safely discarded, or sentences do not implicitly assert anything about their own truth value. Then there are two possible answers to the question of: *"What would be the expectation of the brain with regard to the truth value of a self-referential sentence?"* If Buridan was right, then the answer would be that the expectation of the brain should be *"truth"*, while if he was wrong, the answer would be *"there is no expectation."*

Our results show that there are expectations of the brain with regard to the truth values of self-referential sentences. The Liar sentence fails those expectations, while the Truthteller sentence abides by them. This provides support for Buridan's virtual entailment principle. However, a question may be asked: *"what if it is not falsity but paradoxicality that causes the effort."* Paradoxicality is a concept that arises during the process of reasoning. One stage of reasoning contradicts the next and so forth, leading to a loop. We have studied an earlier stage of thought: meaning comprehension. Brain's effort at that stage can be caused only by an unknown or unexpected word. The word *"false"* is obviously known to everyone; therefore, it is only unexpectedness that could have caused the effort that we observed in our experiment. Because of that, we believe that the best explanation of our results is to interpret that the participants' brains expected the word *"true"* at the ends of self-referential sentences.

Because we need to accept relativism to study human perceptions, it means that our results are of little relevance for deflationist or substantivist theories of truth (for an overview see Wyatt and Lynch 2016). Our experiment does not test if humans perceive an underlying property *truth* that spans across all true sentences. In principle, we showed that integrating the meaning of the Liar sentence is effortful for the brain, just like integrating the meaning of false sentences is. However, this approach works only for sentences that are obviously true or false for the agent that perceives them. If we taught a child from young age that the Sun is a planet, then the brain of that child will react with effort when reading: "The sun is a star." It is possible to be misled and believe in a falsehood. As a result, that falsehood will not cause the N400, but will still be false in the sense of defining truth by correspondence or assertibility.

If our results are built on the relativistic grounds, then it might appear that they should contribute to the contextualist solution of the Liar paradox (Sagi 2016). Unfortunately, a crucial difference separates our paradigm from the contextualist interpretation of the Liar. Contextualists argue that the truth value of sentences depends on the context of their use and assessment. They outline several different steps in the reasoning when an agent determines the truth value of the Liar sentence. Paradox is solved when one accepts that pre-

vious steps in the reasoning, affect the context in the steps that follow. As a result, the Liar sentence has different truth values at different stages of the reasoning (Glanzberg 2001, Glanzberg 2004, Simmons 1993, Simmons 2015). Therefore, one could ask, which stage of the reasoning did we study in our experiment? Here lies the difference between our approach and contextualism. In our experiment we studied the stage of sentence comprehension (i.e. semantic integration of the meaning of a sentence). This places our context even before contextualists formulate the first stage in their reasoning. The first stage in the contextualist approach is constituted by the full, comprehended Liar sentence and followed by conscious reasoning that utilises new, full sentences of natural language. At the moment that we studied (400ms from the onset of the last word in a sentence), there was too little time for conscious thoughts to happen already. These can first appear a few hundred milliseconds later. As a result, we demonstrated that human brain reacts to the Liar sentence like to false sentences at a stage when the contextualist approach does not yet start considering the truth value of the Liar.

Above and beyond the contributions of this experiment to the theories of truth, it also has important implications for the contentual logic. In Fregean logic every sentence can be ascribed one of two logical values of truth or falsehood. This principle generates a number of non-intuitive consequences called "the paradoxes of material implication." One of the most important ones is the tautology of: $(\alpha \rightarrow \beta) \vee (\beta \rightarrow \alpha)$. Because humans think with contents of sentences, it appears that it is not true that for any two sentences at least one follows from another. However, it is a tautology of logical values which merely expresses the truth that: $(1 \rightarrow 0) \vee (0 \rightarrow 1)$. In logic with content implication, alternative: $(\alpha : \beta) \vee (\beta : \alpha)$ is not a tautology. It is not the case that for every two sentences at least one says what is said by the other. In that sense, contentual logic may be closer to how human cognition operates.

We show that the conclusions of non-Fregean logics regarding the Liar paradox coincide with the human comprehension of language. In non-Fregean logics the Liar sentence turns out to be contradictory, which means that the sentence is false and its negation is true. As it turned out, the logic defined on the language extended with the content implication is non-Fregean, just like the Suszko's Sentential Calculus with Identity. As a result, in any logic involving the content implication, the Liar sentence is predicted to be false. Furthermore, when determining the truth value of sentences, contentual logic places heavy emphasis on the agent who produces or perceives them. Because of that, contentual logic also predicts that humans should understand the Liar sentence as false. We provide evidence that this is indeed the case. Perception of sentences from the perspective of their content rather than reducing them to logical values, on one hand, solves the Liar's antinomy, and on the other hand, it is consistent with our cognition.

Acknowledgements

We would like to thank Professor Hanna Bednarek (Department of Cognitive Psychology, SWPS University, Warsaw) who helped us with obtaining the approval of the Ethical Committee for the experiment, and Professor Aneta Brzezicka (Department of Psychophysiology of Cognitive Processes, SWPS University, Warsaw) who allowed us to use the EEG laboratory.

References

1. Benétreau-Dupin, Y. (2015). Buridan's Solution to the Liar Paradox. *History and Philosophy of Logic, 36*(1), 18-28.
2. Bloom, S. L., & Suszko, R. (1972). Investigations into the sentential calculus with identity. *Notre Dame Journal of Formal Logic, 13*(3), 289-308.
3. Bloom, S. L., & Suszko, R. (1975). Ultraproducts of SCI Models. *Bulletin of the Section of Logic, 4*(1), 9-12.
4. Buridan, J. (2001). *Summulae de Dialectica*. trans. G. Klima. New Haven, CT: Yale University Press.
5. Curry, H. B. (1942). The inconsistency of certain formal logics. *The Journal of Symbolic Logic, 7*(3), 115-117.
6. Delorme, A., & Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of neuroscience methods, 134*(1), 9-21.
7. Descartes, R. (1984). Fourth meditation. In J. Cottingham, R. Stoothoff, & D. Murdoch (Eds. and Trans.), textitThe philosophical writings of Descartes. Cambridge, England: Cambridge University Press. (Original work published 1641)
8. Dudschig, C., Maienborn, C. & Kaup, B. (2016). Is there a difference between stripy journeys and stripy ladybirds? The N400 response to semantic and world-knowledge violations during sentence processing. *Brain and cognition, 103*, 38-49.
9. Gilbert, D. T., Krull, D. S. & Malone, P. S. (1990). Unbelieving the unbelievable: Some problems in the rejection of false information. *Journal of personality and social psychology, 59*(4), s. 601-613.
10. Gilbert, D. T., Tafarodi, R. W. & Malone, P. S. (1993). You can't not believe everything you read. *Journal of personality and social psychology, 65*(2), 221-233.
11. Glanzberg, M. (2001). The liar in context. *Philosophical Studies, 103*(3), 217–251.
12. Glanzberg, M. (2004). A contextual-hierarchical approach to truth and the liar paradox. *Journal of Philosophical Logic, 33*(1), 27–88.
13. Hagoort, P., Hald, L., Bastiaansen, M. & Petersson, K. M. (2004). Integration of word meaning and world knowledge in language comprehension. *Science, 304*(5669), 438-441.
14. Hughes, G. E. (1985). *John Buridan on Self-Reference: Chapter Eight of Buridan's Sophismata: With a Translation, an Introduction, and a Philosophical Commentary*. Cambridge: Cambridge University Press.
15. Hyvärinen, A., & Oja, E. (2000). Independent component analysis: algorithms and applications. *Neural networks, 13*(4-5), 411-430.
16. Klima, G. (2018). The Medieval Liar. *Speculum, 93*(1), 121-131.
17. Kripke, S. (1975). Outline of a theory of truth. *Journal of philosophy, 72*(19), 690-716.
18. Lefebvre, N. & Schelein, M., The Liar Lied. *Philosophy Now, 51*
19. Lopez-Calderon, J., & Luck, S. J. (2014). ERPLAB: an open-source toolbox for the analysis of event-related potentials. *Frontiers in human neuroscience, 8*, 213.
20. Łukowski, P. (2011). *Paradoxes*. New York: Springer Science & Business Media.
21. MacFarlane, J. (2014). *Assessment sensitivity: Relative truth and its applications*. OXford: Oxford University Press.
22. Martin, R. L. (1984). *Recent essays on truth and the liar paradox*. Oxford: Oxford University Press.

23. Marques, J. F., Canessa, N., Cappa, S. (2009). Neural differences in the processing of true and false sentences: Insights into the nature of 'truth'in language comprehension. *Cortex, 45*(6), 759-768.
24. Neyman, J., & Pearson, E. S. (1933). IX. On the problem of the most efficient tests of statistical hypotheses. *Phil. Trans. R. Soc. Lond. A, 231*(694-706), 289-337.
25. Nieuwland, M. S., & Kuperberg, G. R. (2008). When the truth is not too hard to handle: An event-related potential study on the pragmatics of negation. *Psychological Science, 19*(12), 1213-1218.
26. G. Priest, R. Routley, & J. Norman (Eds.) (1989). *Paraconsistent Logic: essays on the inconsistent*. Munich: Philosophia Verlag.
27. Prior, A. N. (1961). On a family of paradoxes. *Notre Dame Journal of Formal Logic, 2*(1), 16-32.
28. Rahman, S., Tulenheimo, T. & Genot, E. (Eds.). (2008). *Unity, Truth and the Liar: The Modern Relevance of Medieval Solutions to the Liar Paradox*. New York: Springer Science & Business Media.
29. Read, S. (2002). The liar paradox from John Buridan back to Thomas Bradwardine. *Vivarium, 40*(2), 189-218.
30. Read, S. (2006). Symmetry and paradox. *History and Philosophy of Logic, 27*(4), 307-318.
31. Rescher, N. (1968). Many-valued logic. In N. Rescher, *Topics in Philosophical Logic* (54-125). Dordrecht: Springer.
32. Sagi, G. (2017). Contextualism, Relativism and the Liar. *Erkenntnis, 82*(4), 913-928.
33. Sauerland, U., & Stateva, P. (Eds.). (2007). *Presupposition and implicature in compositional semantics*. New York: Springer.
34. Simmons, K. (1993). *Universality and the liar: An essay on truth and the diagonal argument*. Cambridge University Press.
35. Simmons, K. (2015). Paradox, repetition, revenge. *Topoi, 34*(1), 121–131.
36. Spinoza, B. (1982). *The ethics and selected letters*. (Trans. S. Shirley). Indianapolis, IN: Hackett. (Original work published 1677)
37. Stigler, S. (2008). Fisher and the 5% level. *Chance, 21*(4), 12-12.
38. Tarski, A. (1933). *Pojecie prawdy w jezykach nauk dedukcyjnych* (The concept of truth in languages of deductive sciences). Warszawa: Towarzystwo Naukowe Warszawskie.
39. Uckelman, S. L. (2012). Arthur Prior and medieval logic. Synthese, 188(3), 349-366.
40. Wiswede, D., Koranyi, N., Müller, F., Langner, O. & Rothermund, K. (2012). Validating the truth of propositions: Behavioral and ERP indicators of truth evaluation processes. *Social cognitive and affective neuroscience, 8*(6), s. 647-653.
41. Woleński, J. (1993). Samozwrotność i odrzucanie (Self reference and rejecting). *Filozofia Nauki, 1*(1), 89-102.
42. Wyatt, J., & Lynch, M. (2016). From one to many: recent work on truth. *American Philosophical Quarterly 53*(4):323-340

Appendix A ERP recording

Recording was performed with a 64-channel H_2O-cap, Electrical Geodesics, Inc. system with Cz as the reference electrode. Signal was amplified with a sampling rate of 500 Hz and stored using Net Station software (Electrical Geodesics, Inc.). Pre-processing was performed with the EEGLab software (Delorme and Makeig 2004) with ERPLab plugin (Lopez-Calderon and Luck 2014). Before analysis, a high-pass filter (1 Hz to remove signal drift) and a notch filter (50 Hz to remove powerline noise) were applied. Independent Component Analysis (ICA) was used to remove eye artifacts and other high frequency noise (Hyvärinen and Oja 2000). Grand-average ERPs were created separately for every sentence type across all participants. Time window of 300 ms to 500 ms was used in the statistical analysis.

Table 2 Descriptive statistics of the study variables and tests of the normality of distribution.

Variable name	Mean (μV)	SE	Shapiro-Wilk W	p
N400 for Liar paradox	-0.63	0.18	0.96	0.40
N400 for Truthteller	-0.30	0.16	0.98	0.92
N400 for self-referential falsehood	-0.63	0.18	0.97	0.49
N400 for self-referential truth	-0.16	0.18	0.97	0.61
N400 for normal falsehood	-0.52	0.18	0.98	0.85
N400 for normal truth	-0.17	0.15	0.98	0.79

SE - standard error of the mean

Table 3 Repeated measures ANOVA results for comparing N400 of different types of sentences.

Compared variables	F	df	p	η^2
Liar paradox vs. Self-referential truth	8.0	1, 29	0.008	0.22
Liar paradox vs. Self-referential falsehood	0.0	1, 29	0.98	0.00
Truthteller vs. Self-referential truth	0.8	1, 29	0.38	0.03
Truthteller vs. Self-referential falsehood	4.5	1, 29	0.042	0.24
Self-referential falsehood vs. Self-referential truth	6.9	1, 29	0.013	0.19
Normal truth vs. Normal falsehood	8.0	1, 29	0.008	0.22

Appendix B Visual inspection

Largest effects for self-referential sentences were found on frontal and midline electrodes (Fp2, AFz, Fz, Fp1, AF3, F5). For normal sentences largest effects were found on right frontal electrodes (Fp2, F8). For paradoxical sentences, the negative deflection marking the beginning of the N400 component started around 100ms after the onset of the target word. In false sentences (both self-referential and normal) this latency was longer and negative deflection started around 190ms. Semantic errors (both self-referential and normal) showed extremely higher amplitude of the P2 complex than other types of sentences, with negative deflection starting at the same time as false sentences (190ms), but lasting much longer and peaking around 600ms. True sentences (both self-referential and normal) showed no presence of the N1-P2 complex, instead steadily dropping until 300ms with the lowest amplitude of all types of sentences. The highest N400 amplitude was elicited by paradoxical sentences, second highest by false sentences. True sentences and the Truthteller sentences did not differ in N400 amplitude, with both being lower than paradoxes and falsehoods. Because semantic errors showed highly different latency of the negative deflection it is not possible to accurately assess their N400 component. Additionally, following the N400 time window there was a more negative waveform for false and paradoxical sentences.

Appendix C Statistical analysis

A statistical analysis (repeated measures ANOVA) was performed to test the hypothesis that paradoxical sentences are processed like false sentences. Re-

peated measures ANOVA was selected due to the fact that the experiment followed a within-design and to compare the N400 for different types of sentences a dependent-samples test was required. In Table 2 we report the mean values of N400 amplitudes for different types of sentences. Because ANOVA requires the variables to be normally distributed, we also report the results of the Shapiro-Wilk test of normality in Table 2. All of the study variables were found to be normally distributed.

The detailed results of the ANOVA analysis are reported in Table 3.

First, to confirm the validity of our signal recording we checked if normal false sentences elicited different N400 than normal true sentences. Consistently with previous literature we have found that normal false sentences elicited higher N400 than normal true sentences.

Next, we performed the analysis for the main hypotheses of the study. It revealed that the N400 component was significantly higher for paradoxes compared to true self-referential sentences and that false self-referential sentences also elicited higher N400 than true self-referential sentences.

We have also checked if the Liar sentences elicited different N400 than false self-referential sentences and found that they did not. In fact, mean N400 for the Liar sentences was identical to that of false self-referential sentences with accuracy to two decimal places (see Table 2).

To make sure that the effect is specific for the Liar sentence itself, and not only due to the fact that it address its own truth value, we also performed analysis for the Truthteller sentences. It revealed that the Truthteller sentences did not elicit N400 different from true self-referential sentences.

Additional, exploratory analysis was performed to check if participants' opinions about the truth value of the Liar sentences and their previous knowledge of it had any impact on their ERPs. 18 participants believed the Liar sentence to be false, 5 to be true and 7 answered "neither". 21 participants indicated that they did not know what the Liar paradox is, while 9 that they did.

There was no effect of the participants' opinion about the truth value of the Liar sentence on their ERPs: $F(2, 27) = 0.8, p = 0.46, \eta^2 = 0.06$.

We also checked if including the previous knowledge of the Liar paradox in the model will diminish the effect of the Liar sentence on N400 (i.e. if excluding the people who knew about the paradox will affect the result). Interaction between previous knowledge and the effect of the Liar sentence was not significant: $F(1, 28) = 2.3, p = 0.14, \eta^2 = 0.08$, while the main effect of the Liar remained significant: $F(1, 28) = 4.2, p = 0.05, \eta^2 = 0.13$, even though the sample size got reduced by 9 people (almost a third of the original sample). There is a visible trend in people who had previous knowledge of the Liar paradox to exhibit lower N400 component in response to the Liar sentence, although it is not statistically significant. This trend is present only for the Liar sentence, but not for other types of sentences (see Figure 5).

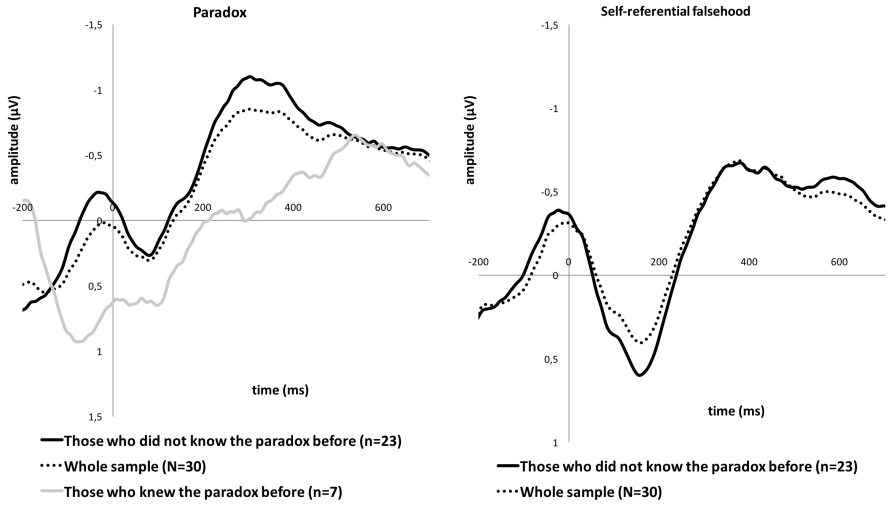

Fig. 5 Event-related potentials on representative electrodes for self-referential sentences in the whole sample and in the subset who did not know the Liar paradox before. Signal was exponentially smoothed for visualization. Smoothing constant = 0.98