Published in Martin Blicha and Igor Sedlár (eds.). *The Logica Yearbook 2020*. London: College Publications, 2021, pp. 183-199. ISBN 978-1-84890-376-0.

A Note on Paradoxical Propositions from an Inferential Point of View

IVO PEZLAR[1]

Abstract: In a recent paper by Tranchini (2019), an introduction rule for the paradoxical proposition ρ^* that can be simultaneously proven and disproven is discussed. This rule is formalized in Martin-Löf's constructive type theory (CTT) and supplemented with an inferential explanation in the style of Brouwer-Heyting-Kolmogorov semantics. I will, however, argue that the provided formalization is problematic because what is paradoxical about ρ^* from the viewpoint of CTT is not its provability, but whether it is a proposition at all.

Keywords: proof-theoretic semantics, constructive type theory, paradox, inductive definitions, Martin-Löf

1 Introduction

How do we define the meaning of logical constants? What does, e.g., the conjunction $\wedge$ mean? The standard answer put forward by the inferentialist (proof-theoretic) tradition is relatively simple: the meaning of logical constants within a certain natural deduction system is specified by introduction rules. These rules should effectively work as " 'definitions' of the symbols concerned" (Gentzen, 1969, p. 80, English translation). For example, the meaning constituting introduction rule for the conjunction $\wedge$ can be schematized as follows:

$$\frac{A \qquad B}{A \wedge B} \wedge\mathrm{I}$$

with the assumption that A and B are true, i.e., proven, propositions. It tells us that if we want to prove the proposition $A \wedge B$, first we have to prove the propositions A and B. In other words, the proof of $A \wedge B$ consists of a pair

[1] Work on this paper was supported by Grant no. 19-12420S from the Czech Science Foundation, GA ČR. I would also like to thank the anonymous reviewers for their most helpful comments.

of proofs of its conjuncts. Thus, the general idea is that we understand the meaning of $\wedge$ when we can properly use it, which in this case corresponds to the ability to prove the proposition of the form $A \wedge B$.

Where do introduction rules come from? The original Gentzen's set of introduction rules arose from analysing the structure of actual mathematical proofs. He wanted to capture "the forms of deduction used in practice in mathematical proofs" and develop a formal system that would come "as close as possible to actual reasoning" (Gentzen, 1969, p. 68). From this perspective, the origins of introduction rules were purely empirical. Hence, there were no general restrictions on them, they were—or rather should have been—just codifying the actual mathematical practice.

However, considering introduction rules should act as "definitions" of the symbols appearing in their conclusions, it seems reasonable to assume that the symbols to be defined should not appear among the premises of the corresponding rules to avoid circularity.[2] For example, assume that the introduction rule for conjunction would look as follows:

$$\frac{A \wedge B}{A \wedge B}\ \wedge\mathrm{I}'$$

It would be difficult to see in what sense it constitutes or illuminates the meaning of $\wedge$.

Furthermore, if these "definitions" provided by introduction rules are to be of any practical value, we have to know how to use them. In natural deduction systems, this is a task for elimination rules, whose role is, simply put, to enact those definitions. In Gentzen's words: "[elimination rules] are no more, in the final analysis, than the consequences of these definitions [i.e., of introduction rules]" (Gentzen, 1969, p. 80). For example, the elimination rules for conjunction are as follows:

$$\frac{A \wedge B}{A}\ \wedge\mathrm{E}_1 \qquad\qquad \frac{A \wedge B}{B}\ \wedge\mathrm{E}_2$$

It seems unproblematic that if $A \wedge B$ was "defined" using A and B, we should be able to unpack this definition and get back its constituents, i.e., A and B in this case.

The observation that we should not be able to infer from a derived proposition more (or less) than what went into its derivation is crucial. It was this

[2]However, as we will see later, this is not always so straightforward, especially in the case of inductive definitions.

general concept—that later become known as the inversion principle (see Lorenzen, 1955; Prawitz, 1965) or harmony (see Dummett, 1991; Tennant, 1978)—that was violated by the famous counterexample by Prior (1960) to the idea that introduction and elimination rules alone can determine the meaning of logical constants. He proposed a new logical constant tonk governed by the following introduction and elimination rules:

$$\frac{A}{A \ \mathtt{tonk} \ B}\,\mathtt{tonk}\mathrm{I} \qquad \frac{A \ \mathtt{tonk} \ B}{B}\,\mathtt{tonk}\mathrm{E}$$

With these rules it is easy to derive the paradoxical conclusion that $\neg A$ is true assuming that A is true:

$$\cfrac{\cfrac{A}{A \ \mathtt{tonk} \ \neg A}\,\mathtt{tonk}\mathrm{I}}{\neg A}\,\mathtt{tonk}\mathrm{E}$$

What went wrong? It is the elimination rule that causes the paradoxical behaviour. Specifically, the elimination rule is not sanctioned by the corresponding introduction rule. As was said, elimination rules should not go beyond what introduction rules stipulate. In this case, it is the derivation of $A \ \mathtt{tonk} \ B$ from A. And since no B went into deriving $A \ \mathtt{tonk} \ B$, we should not be able to derive B back from it.[3]

However, as was already observed by Prawitz (1965, Appendix B), there are scenarios in which introduction and elimination rules are harmonious, yet paradoxical behaviour still arises. In these cases, the culprit is not the elimination rules as was the case with tonk, but the introduction rules.[4]

It is these problematic introduction rules, namely those that exhibit paradoxical behaviour due to some form of circularity, that will be the main topic of this paper. Specifically, I will examine an introduction rule discussed by Tranchini (2019) determining the meaning of the paradoxical proposition ρ^* that can be simultaneously proven and disproven, i.e., we can have proofs for both ρ^* and its negation. In the same paper, this rule is then formalized in the framework of Martin-Löf's constructive type theory (CTT) and supplied with a corresponding clause to the Brouwer-Heyting-Kolmogorov (BHK)

[3]See (Tranchini, 2014) for a more thorough discussion of tonk and its difference from other paradoxical connectives.

[4]I would like to thank the reviewer for this remark. See also (Schroeder-Heister, 2012) for a discussion of paradoxical behaviour in the sequent calculus setting, i.e., in an environment without introduction and elimination rules but with left and right rules.

semantics. I will, however, argue that the provided formalization is problematic because what is paradoxical about ρ^* from the viewpoint of CTT is not its provability, but whether it is a proposition at all.

The introduction and elimination rules and BHK clause for the paradoxical proposition ρ^* proposed by Tranchini (2019) are based on the introduction and elimination rules and BHK clause for implication $\supset$. So we begin by examining the latter, then we discuss the former.

2 Implication

According to the BHK semantics, the proof, and hence the meaning of the proposition $A \supset B$ consists of a method (procedure, function, program) which takes any proof of A and returns a proof of B. The standard introduction and elimination rules are as follows:

$$\frac{\begin{array}{c}[A]\\ B\end{array}}{A \supset B}\ \supset\text{I} \qquad\qquad \frac{A \supset B \qquad A}{B}\ \supset\text{E}$$

The introduction rule tells us that if we want to prove proposition $A \supset B$, we first have to be able to derive proposition B from assumption A which should then be discharged.[5] In other words, it tells us how to construct (canonical) proofs of the proposition $A \supset B$. The elimination rule then tells us how we can use this proposition in proofs: if we derive $A \supset B$ together with A, we can then proceed to B alone.

Note that if we apply the $\supset$E rule immediately after the $\supset$I rule, i.e., construct $A \supset B$ and then remove it right away, we are making an unnecessary detour in a derivation. To get rid of these detours we use the following reduction meta rule (see Prawitz, 1965):

$$\frac{\dfrac{\begin{array}{c}[A]^n\\ \mathcal{D}\\ B\end{array}}{A \supset B}\ \supset\text{I}^n \qquad \begin{array}{c}\mathcal{D}'\\ A\end{array}}{B} \quad\Rightarrow\quad \begin{array}{c}\mathcal{D}'\\ [A]\\ \mathcal{D}\\ B\end{array}$$

If a closed derivation, i.e., a derivation with no open assumptions, contains no detours, it is said to be in normal form.

[5] Hence, the assumption A is essentially just a placeholder to be withdrawn. For an alternative approach to assumptions, see (Pezlar, 2020).

Utilizing the propositions-as-types principle[6], which is fully adopted by Martin-Löf's constructive type theory (CTT), we can make the BHK clause as well as the rules for implication more explicit and precise:

$$\dfrac{\begin{matrix}[x : A] \\ b(x) : B\end{matrix}}{\lambda x.b(x) : A \supset B}\,\supset\mathrm{I} \qquad\qquad \dfrac{c : A \supset B \qquad a : A}{\mathtt{app}(c, a) : B}\,\supset\mathrm{E}$$

assuming that A : *prop*, B : *prop*, and $A \supset B$: *prop*, i.e., that A, B, and $A \supset B$ are propositions. The meta rule for detour reduction can then be captured as the following computation rule (also known as reduction rule or equality rule):

$$\dfrac{\begin{matrix}[x : A] \\ t(x) : B\end{matrix} \qquad s : A}{\mathtt{app}(\lambda x.t(x), s) = t(s/x) : B}\,\supset\mathrm{C}$$

where $t(s/x)$ is the result of substituting s for x in t. Informally, the rule states that a derivation with a detour is equal to the derivation we obtain by removing this detour.

In what sense are these rules more explicit and precise? Regarding the explicitness, note that the premises and conclusions of these rules are no longer propositions but judgments of the form $a : A$ which can be read as "a is a proof of A". Hence, the proofs themselves are internalized in the object language and coded as terms. As for the precision, note that the informal statement of the corresponding BHK clause "a proof of $A \supset B$ consists of a method that takes any proof of A and returns a proof of B" is made more exact by the judgment $\lambda x.b(x) : A \supset B$ where the unspecified notion of a method is replaced by a specific lambda term, namely abstraction.

It is important to mention that from the perspective of CTT, the function $\lambda x.b(x)$ appearing in the conclusion of $\supset$I is a function in a secondary sense, the more basic notion of a function appears in the premise of this rule, i.e., it is captured by the hypothetical derivation of $b(x) : B$ under the assumption $x : A$ (see, e.g., Klev, 2019a). We can liken this difference to Frege's distinction between functions as course-of-values and functions as unsaturated entities, respectively (see Frege, 1893).

[6] Also known as the Curry-Howard correspondence (see Curry & Feys, 1958; Howard, 1980).

One last note. So far we have presupposed that $A \supset B$ is a proposition, assuming A and B are propositions. In CTT, however, this is a judgment that can and should be demonstrated as well by using a special kind of rules called formation rules. These rules tell us how to form new propositions from other propositions. For example, $A \supset B$ receives the following formation rule:[7]

$$\frac{A : prop \qquad B : prop}{A \supset B : prop} \supset\text{F}$$

Note that this rule tells us how to form the proposition $A \supset B$, i.e., how to derive the judgment $A \supset B : prop$. However, the rule itself requires further justification. Generally, in CTT, we can judge that A is a proposition if we know what counts as a canonical proof of A, i.e., if we can recognize a canonical proof of A when we are presented with one (the same goes for equal canonical proofs). And to tell us what counts as canonical proofs is the purpose of the introduction rules. Therefore, formation rules are justified by the corresponding introduction rules. Thus, we can judge that $A \supset B : prop$ since we know (via the $\supset$I rule) what should the canonical proofs of $A \supset B$ look like.

Consequently, this means that the rule $\supset$I takes, if we want to be fully explicit, three premises, including those of the corresponding formation rules. The rule then should look as follows:

$$\frac{A : prop \qquad B : prop \qquad \begin{matrix}[x : A] \\ b(x) : B\end{matrix}}{\lambda x.b(x) : A \supset B} \supset\text{I}'$$

Note In CTT, there are four basic kinds of rules: introduction rules, elimination rules, formation rules, and computation rules. Introduction rules are considered self-justifying, elimination rules correspond to introduction rules and are justified by computation rules (analogous to Prawitz's reduction rules, see Prawitz, 1965), formation rules are justified by introduction rules, and computation rules relate elimination rules to introduction rules. For more, see (Martin-Löf, 1984).

Now, let us finally proceed to the paradoxical proposition ρ^*.

[7] I am omitting the variant for showing how to form equal implicational propositions.

3 Paradoxical proposition ρ^*

To incorporate the paradoxical proposition ρ^* that can be simultaneously proven and disproven, Tranchini (2019) suggests the following extension to the BHK semantics. The informal clause explaining the corresponding proof condition for ρ^* goes as follows, where $\bot$ denotes absurdity:

> a proof of ρ^* is the result of applying a self-referential abstraction-like operation to a function (as an unsaturated entity) from proofs of ρ^* to proofs of $\bot$. The result of this operation are objects whose nature is similar to that of the functions as courses-of-value that constitute proofs of sentences of the form $A \supset B$, with the crucial difference that proofs of ρ^* take proofs of ρ^* as arguments and yield proofs of $\bot$ as values. (Tranchini, 2019, p. 601)

The clause for the paradoxical proposition ρ^* is given formalization in the framework of Martin-Löf's constructive type theory (CTT). Specifically, its inferential behaviour is specified by the following introduction and elimination rules (see also Read, 2010):

$$\frac{\begin{array}{c}[x : \rho^*] \\ t(x) : \bot\end{array}}{⅄x.t(x) : \rho^*}\,\rho^*\mathrm{I} \qquad \frac{s : \rho^* \qquad t : \rho^*}{\mathtt{ɐbb}(s, t) : \bot}\,\rho^*\mathrm{E}$$

which are then related by the following computation rule:

$$\frac{\begin{array}{c}[x : \rho^*] \\ t(x) : \bot\end{array} \qquad s : \rho^*}{\mathtt{ɐbb}(⅄x.t(x), s) = t(s/x) : \bot}\,\rho^*\mathrm{C}$$

The ρ^*C rule shows how the function $\mathtt{ɐbb}$, defined by the rule ρ^*E, operates on the canonical proofs of the proposition ρ^* generated by the rule ρ^*I, and thus in turn justifies the ρ^*E rule.

With these rules we can both prove ρ^* and disprove ρ^*, i.e., prove $\neg\rho^*$ understood as $\rho^* \supset \bot$:

$$\frac{\dfrac{[x : \rho^*]^1 \qquad [x : \rho^*]^1}{\mathtt{ɐbb}(x, x) : \bot}\,\rho^*\mathrm{E}}{⅄x.\mathtt{ɐbb}(x, x) : \rho^*}\,\rho^*\mathrm{I}^1 \qquad \frac{\dfrac{[x : \rho^*]^1 \qquad [x : \rho^*]^1}{\mathtt{ɐbb}(x, x) : \bot}\,\rho^*\mathrm{E}}{\lambda x.\mathtt{ɐbb}(x, x) : \rho^* \supset \bot}\,\supset\mathrm{I}^1$$

If we combine these two derivations via the $\supset$E rule, we obtain the following derivation:

$$\dfrac{\dfrac{\dfrac{[x : \rho^*]^1 \qquad [x : \rho^*]^1}{\mathtt{ɐbb}(x,x) : \bot}\,\rho^*\mathrm{E}}{\text{⅄}x.\mathtt{ɐbb}(x,x) : \rho^*}\,\rho^*\mathrm{I}^1 \qquad \dfrac{\dfrac{[x : \rho^*]^1 \qquad [x : \rho^*]^1}{\mathtt{ɐbb}(x,x) : \bot}\,\rho^*\mathrm{E}}{\lambda x.\mathtt{ɐbb}(x,x) : \rho^* \supset \bot}\,{\supset}\mathrm{I}^1}{\mathtt{app}(\lambda x.\mathtt{ɐbb}(x,x), \text{⅄}x.\mathtt{ɐbb}(x,x)) : \bot}\,{\supset}\mathrm{E}$$

Note that there is a redundancy on the right side of the tree: we derived $\rho^* \supset \bot$ via the $\supset$I rule and then immediately eliminated it by an application of the $\supset$E rule.

If we remove this detour using the $\supset$C rule (i.e., essentially compute the term $\mathtt{app}(\lambda x.\mathtt{ɐbb}(x,x), \text{⅄}x.\mathtt{ɐbb}(x,x))$), we get the following derivation:[8]

$$\dfrac{\dfrac{\dfrac{[x : \rho^*]^1 \qquad [x : \rho^*]^1}{\mathtt{ɐbb}(x,x) : \bot}\,\rho^*\mathrm{E}}{\text{⅄}x.\mathtt{ɐbb}(x,x) : \rho^*}\,\rho^*\mathrm{I}^1 \qquad \dfrac{\dfrac{[x : \rho^*]^1 \qquad [x : \rho^*]^1}{\mathtt{ɐbb}(x,x) : \bot}\,\rho^*\mathrm{E}}{\text{⅄}x.\mathtt{ɐbb}(x,x) : \rho^*}\,\rho^*\mathrm{I}^1}{\mathtt{ɐbb}(\text{⅄}x.\mathtt{ɐbb}(x,x), \text{⅄}x.\mathtt{ɐbb}(x,x)) : \bot}\,\rho^*\mathrm{E}$$

with yet another detour. But if we try to remove it, we discover that this derivation reduces to itself via the ρ^*C rule, and thus we get caught in what Neil Tennant called a loop: "the normalisation sequence never terminating with a proof in normal form" (Tennant, 1982, p. 270).

As was already mentioned before, note that the rule ρ^*I is essentially assembled as a self-referential variant of the implication introduction rule $\supset$I, which, as we discussed above, takes, in its fully explicit version, three premises, not just one. So, analogously, the fully revealed version of the ρ^*I rule for the proposition ρ^* should be:

$$\dfrac{\rho^* : prop \qquad \bot : prop \qquad \begin{matrix}[x : \rho^*] \\ t(x) : \bot\end{matrix}}{\text{⅄}x.t(x) : \rho^*}\,\rho^*\mathrm{I}'$$

Now, let us examine more closely the object $\text{⅄}x.t(x)$ of type ρ^*, i.e., the object constructed by the ρ^*I rule. Analogously to $\lambda x.b(x)$, it is supposed to be a coding (a name) of the function $t(x)$. Note, however, that the domain of $t(x)$ is also ρ^*, i.e., $\text{⅄}x.t(x)$ itself belongs to the domain ρ^* of the function

[8] Alternatively, we could construct this derivation directly from the two copies of the initial closed derivation of ρ^*, as does Tranchini (2019).

$t(x)$ it is supposed to be coding. But that is a problem: we are forming a canonical object of type ρ^* from a function whose domain is ρ^* itself. In other words, we are generating an object of type ρ^* from itself (see Dyckhoff, 2016 for analogous observations).[9]

But is this really problematic in general? For example, the type $\mathbb{N}$ of natural numbers seems to be defined also with some degree of circularity but everything works just fine. Specifically, it has the following two introduction rules:

$$\frac{}{0 : \mathbb{N}}\ \mathbb{N}\mathrm{I}_1 \qquad\qquad \frac{n : \mathbb{N}}{succ(n) : \mathbb{N}}\ \mathbb{N}\mathrm{I}_2$$

The first rule simply stipulates that 0 is a natural number, so there is no issue. However, in the case of the second rule it seems like we are trying to generate an object of type $\mathbb{N}$ from itself: note that the premise of the successor rule $\mathbb{N}\mathrm{I}_2$ seems to presuppose that we already understand what it means for some n to be a natural number.

Why is this case unproblematic as opposed to $\rho^*\mathrm{I}$? Is it perhaps because with the type $\mathbb{N}$ we have the base object $0 : \mathbb{N}$ which is missing in the case of the type ρ^*? Unfortunately no, because then we would have to conclude that the type of well-founded trees (W-types) has also problematic introduction rules, since they also do not have such base cases.[10]

The answer to this question can be found, if we examine the premises of the involved rules. Specifically, note that the rule ρ^* tells us that we can construct a canonical proof of ρ^* assuming we have a function that takes an arbitrary proof of ρ^* and returns a proof of $\bot$ (recall the BHK explanation of implication).

In contrast, the premise of the successor rule for $\mathbb{N}$ does require us to be in a possession of an arbitrary object of this type, we just need an object of this type. Or as Dyckhoff (2016, p. 82) put it: "we don't need to grasp *all* elements of $\mathbb{N}$ to construct a canonical element by means of the rule, just one of them, namely n."

To make these observations more general and precise, we can borrow a few notions from the literature on inductive types and then carry them over to the logical side in accordance with the Curry-Howard correspondence.[11]

[9] I thank one of the reviewers for pointing this out to me and invite the reader to consult Dyckhoff's paper as well, specifically pp. 81–83. Furthermore, the elimination rule $\rho^*\mathrm{E}$ is also of some interest, however, in this paper I will be focused primarily on introduction rules.

[10] I thank an anonymous reviewer for this remark. Although W-types would deserve a closer examination, they are beyond the scope of this paper.

[11] As noted, e.g., by Dyckhoff (2016), whose simplified presentation we follow below.

Inductive types are often formally treated as the least fixed point of an operator Φ defined via type variables $X, Y, \ldots$, constants, and type constructors $+$ (addition), $\times$ (product), and $\rightarrow$ (function space). For example, the type of natural numbers $\mathbb{N}$ can be generated by the definition $\Phi(X) = X + 1$.[12]

Furthermore, a definition of a unary operator Φ is said to be *positive* if and only if only occurrences of X in it are positive.

1. An occurrence of X in X is *positive*.

2. An occurrence of X in $A \rightarrow B$ is *positive* if and only if it is (i) a positive occurrence in B or (ii) a negative occurrence in A

3. An occurrence of X in $A \rightarrow B$ is *negative* if and only if it is (i) a negative occurrence in B or (ii) a positive occurrence in A

4. An occurrence of X in $A + B$ and $A \times B$ is *positive* if and only if it is a positive occurrence in A or B

5. An occurrence of X in $A + B$ and $A \times B$ is *negative* if and only if it is a negative occurrence in A or B

A definition of a type as the least fixed point of an operator can then be said to be *positive* if and only if the operator definition is positive. Furthermore, a definition can be said to be *strictly positive*, if only occurrences of the type variable X in the definition are strictly positive.

6. An occurrence of X in X is *strictly positive*.

7. An occurrence of X in $A \rightarrow B$ is *strictly positive* if and only if it is a strictly positive occurrence in B

8. An occurrence of X in $A + B$ and $A \times B$ is *strictly positive* if and only if it is a strictly positive occurrence in A or B

For example, in $A \rightarrow B$, B occurs positively and A occurs negatively. Also note that if X does not occur positively in A then either X does not occur in A or X occurs negatively in A.

Now, the type ρ^* can be then defined as the least fixed point of the operator $\Phi(X) = X \rightarrow \bot$ (see Dyckhoff, 2016, p. 83). Note that this is not a positive definition, since X has a negative occurrence in the definition (see the clause 3

[12] See, e.g., (Dybjer, 1997), (Mendler, 1987).

above). Definitions with negative occurrences such as this are generally avoided (so called "strict positivity condition") for the unsurprising reason that they allow us to construct looping terms, analogous to our non-normalizable looping proofs (see, e.g., Bertot & Castéran, 2004; Chlipala, 2013).

Intuitively, we can think of these type definitions as corresponding to introduction rules under the Curry-Howard correspondence.[13] Analogously, we can also carry over the notions of positive/negative occurrence. For example, in the case of ρ^*, the negative occurrence of ρ^* in the corresponding type definition then coincides with the fact that on the logical side ρ^* appeared as an assumption. Similarly, we can also adopt the notions of (strictly) positive/negation definitions, e.g., we can say that an introduction rule for a proposition A is strictly positive if and only if A does not appear among its premises as an assumption/antecedent.[14]

Now, if we return to the difference between $\rho^*\mathrm{I}$ and $\mathbb{N}\mathrm{I}_2$, we can then say that the reason why the former is problematic, but the latter is not, is because ρ^* occurs negatively in a premise of $\rho^*\mathrm{I}$, which is not the case for $\mathbb{N}$ in $\mathbb{N}\mathrm{I}_2$. Thus, we can say that $\rho^*\mathrm{I}$ is a negative introduction rule.

So, to conclude, the problem with $\rho^*\mathrm{I}$ is not just that ρ^* itself appears in the premise (as we have seen, e.g., $\mathbb{N}$ also appears in the premise of the corresponding rule $\mathbb{N}\mathrm{I}_2$ but causes no issues) but that it occurs negatively in the premise. Why is this problematic from the viewpoint of CTT? Recall that the premise of $\rho^*\mathrm{I}$ is a hypothetical judgment stating that $t(x)$ is a function from ρ^* to $\perp$. In order to fully understand this function, however, we have to understand what an arbitrary proof of ρ^* is. But to achieve this, we first need to understand what a canonical proof of ρ^* is. And to understand this, we need to understand the corresponding introduction rule. Thus, understanding the premise of this rule presupposes that we already understand the rule as a whole.[15] More generally put, the rule invites us to assume we know something (what does the canonical proof object of ρ^* look like) which is unknowable at that point. For these reasons, the rule $\rho^*\mathrm{I}$ cannot be considered as properly constituting the meaning of ρ^*.

Furthermore, recall that in CTT, formation rules are justified by introduction rules. In practice, this means that to be able to make the judgment

[13] The type constructor $\rightarrow$ can be roughly understood as corresponding to the implication operator $\supset$ on the logical side.

[14] As observed by Klev (2019b), negative occurrences of propositions in their own introduction rules are already banned implicitly by Martin-Löf (1971), explicitly by Dybjer (1994).

[15] For an analogous observation, see also (Klev, 2019b). I thank one of the reviewers for pointing this out.

that ρ^* is a proposition, i.e., $\rho^* : prop$, we first have to be able to show how its canonical proofs are constructed. But, as we just showed, this cannot be done, since the rule ρ^*I fails in its task to explain properly the meaning of ρ^*, i.e., of its canonical proofs. Consequently, we are not in a position to justifiably make the judgment $\rho^* : prop$.

Therefore, if there is a paradox from the viewpoint of CTT, it is rather about the formability of ρ^*. In other words, what can perhaps be viewed as paradoxical is the justifiability and explainability of the judgment $\rho^* : prop$. Utilizing the rules above we can provide some form of an "explanation" but simultaneously it cannot really be considered as a proper explanation due to its circular nature and the negative occurrence of ρ^* in the premise of its introduction rule can be understood as an indication of this. Thus, what seems to be in question is not the provability of ρ^*, but whether it is a proposition at all.[16]

4 Variants of ρ^*

The other variants of ρ^* discussed by Tranchini (2019) seem to suffer from analogous issues and, in the remaining place, I will try to briefly sketch why. These variants are: 1) a paradoxical proposition ρ with a negative self-reference operator ! and its inverse ¡ and 2) semi-paradoxical propositions σ and τ whose paradoxical nature does not come from self-reference, or negative self-reference, but from their circular meaning-dependencies.

First, we consider the paradoxical proposition ρ with a negative self-reference operator. It is governed by the following introduction and elimination rules:

$$\frac{t : \neg\rho}{!t : \rho}\,\rho\mathrm{I} \qquad \frac{t : \rho}{¡t : \neg\rho}\,\rho\mathrm{E}$$

The corresponding computation rule is as follows:

$$\frac{t : \neg\rho}{¡!t = t : \neg\rho}\,\rho\mathrm{C}$$

Analogously to ρ^*, with these rules we can construct proofs for ρ as well as $\neg\rho$, i.e., $\rho \supset \bot$, and combine them into a looping derivation:

[16] That is, of course, not to say that Russell-like self-referential paradoxes cannot be recreated in CTT at all (see, e.g., Coquand, 1992).

$$\dfrac{\dfrac{\dfrac{\dfrac{[x:\rho]^1}{\text{¡}x:\neg\rho}\,\rho\text{E} \qquad [x:\rho]^1}{\mathtt{app}(\text{¡}x,x):\bot}\,\supset\text{E}}{\lambda x.\mathtt{app}(\text{¡}x,x):\neg\rho}\,\supset\text{I}^1 \qquad \dfrac{\dfrac{\dfrac{\dfrac{[x:\rho]^1}{\text{¡}x:\neg\rho}\,\rho\text{E} \qquad [x:\rho]^1}{\mathtt{app}(\text{¡}x,x):\bot}\,\supset\text{E}}{\lambda x.\mathtt{app}(\text{¡}x,x):\neg\rho}\,\supset\text{I}^1}{!\lambda x.\mathtt{app}(\text{¡}x,x):\rho}\,\rho\text{I}}{\mathtt{app}(\lambda x.\mathtt{app}(\text{¡}x,x),!\lambda x.\mathtt{app}(\text{¡}x,x)):\bot}\,\supset\text{E}$$

Now, returning to the rule ρI, if we make it fully explicit, we obtain:

$$\dfrac{\rho : prop \qquad t:\rho\supset\bot}{!t:\rho}\,\rho\text{I}'$$

which commits the same violation as the rule ρ^*I. Simply put, the rule ρI$'$, which should act as a meaning explanation for ρ, presupposes that we already understand ρ. Here again it is indicated by the negative occurrence of ρ in the premise of the ρI rule, analogously to ρ^*I. Consequently, we are not justified in making the judgment $\rho : prop$.

Now, let us consider the case involving the propositions σ and τ with circular meaning-dependencies. They are specified by the following introduction and elimination rules:[17]

$$\dfrac{\begin{matrix}[x:\tau]\\ t(x):\bot\end{matrix}}{\text{⅄}'x.t(x):\sigma}\,\sigma\text{I} \qquad \dfrac{s:\sigma \qquad t:\tau}{\mathtt{ɐbb}'(s,t):\bot}\,\sigma\text{E} \qquad \dfrac{s:\sigma}{!'s:\tau}\,\tau\text{I} \qquad \dfrac{t:\tau}{\text{¡}'t:\sigma}\,\tau\text{E}$$

Although Tranchini does not supplant them with computation rules, I assume they might look as follows:

$$\dfrac{\begin{matrix}[x:\tau]\\ t(x):\bot\end{matrix} \qquad s:\sigma}{\mathtt{ɐbb}'(\text{⅄}'x.t(x),s)=t(s/x):\bot}\,\sigma\text{C} \qquad \dfrac{t:\tau}{\text{¡}'!'t=t:\tau}\,\tau\text{C}$$

Finally, Tranchini hints at the end of his paper that with these rules we can recreate Jourdain's paradox. To test it out, consider the following derivation:

[17]Tranchini (2019) presents these rules without the corresponding proof terms. Here we assume that the operators marked with '$'$' behave analogously to their unmarked variants with the exceptions generated by different typing (e.g., ! expects $\neg\rho$, while !$'$ expects σ).

$$\dfrac{\dfrac{\dfrac{\dfrac{[x:\tau]^1}{\text{¡}'x:\sigma}\,\tau\mathrm{E} \qquad [x:\tau]^1}{\mathtt{\text{ɐdd}}'(\text{¡}'x,x):\bot}\,\sigma\mathrm{E}}{\text{⅄}'x.\mathtt{\text{ɐdd}}'(\text{¡}'x,x):\sigma}\,\sigma\mathrm{I}^1 \qquad \dfrac{\dfrac{\dfrac{\dfrac{[x:\tau]^1}{\text{¡}'x:\sigma}\,\tau\mathrm{E} \qquad [x:\tau]^1}{\mathtt{\text{ɐdd}}'(\text{¡}'x,x):\bot}\,\sigma\mathrm{E}}{\text{⅄}'x.\mathtt{\text{ɐdd}}'(\text{¡}'x,x):\sigma}\,\sigma\mathrm{I}^1}{\text{¡}'\text{⅄}'x.\mathtt{\text{ɐdd}}'(\text{¡}'x,x):\tau}\,\tau\mathrm{I}}{\mathtt{\text{ɐdd}}'(\text{⅄}'x.\mathtt{\text{ɐdd}}'(\text{¡}'x,x),\text{¡}'\text{⅄}'x.\mathtt{\text{ɐdd}}'(\text{¡}'x,x)):\bot}\,\sigma\mathrm{E}$$

Again, if we try to reduce this derivation to a normal form, it enters into a loop (compare this with the derivation of $\mathtt{app}(\lambda x.\mathtt{app}(\mathsf{i}x,x),!\lambda x.\mathtt{app}(\mathsf{i}x,x)):\bot$ from earlier).

Now, let us return to the introduction rules σI and τI. It is clear that the respective introduction rules are dependent on each other. Specifically, σI presupposes that we already understand what τ means and τI presupposes that we already understand what σ means. Again, let us begin by considering their fully explicit versions. We start with the rule σI:

$$\dfrac{\tau:prop \qquad \bot:prop \qquad \begin{matrix}[x:\tau]\\ t(x):\bot\end{matrix}}{\text{⅄}'x.t(x):\sigma}\,\sigma\mathrm{I}'$$

First, note that there is no apparent circularity, as the corresponding type definition $\Phi(X)=Y\to\bot$ suggests. But also note that $\sigma\mathrm{I}'$ is in a way an incompletely specified rule since it refers to τ which we do not yet know to be a proposition. More specifically, we do not yet know how to prove it, i.e., how to construct its canonical proofs. For that, we need to consider an additional introduction rule τI (recall that the meanings of σ and τ should be interdependent). Once again, let us consider its explicit version:

$$\dfrac{\sigma:prop \qquad s:\sigma}{\text{¡}'s:\tau}\,\tau\mathrm{I}'$$

Again, at this stage, still no circularity appears, as the corresponding type definition $\Phi(Y)=X$ shows us. Note, however, that if we put these two definitions together, once again we obtain a negative definition of the form $\Phi(X)=X\to\bot$ (either if we substitute X for Y in the type definition corresponding to $\sigma\mathrm{I}'$ or $Y\to\bot$ for X in the type definition corresponding to $\tau\mathrm{I}'$). Thus, a complete specification of these introduction rules seems to come at a price of negativity.

Note The Liar-like paradoxes considered in this paper can be generalized to a Curry-like paradox. For example, Read (2010) considers a proof-theoretic variant of a Curry paradox specified by the following introduction and elimination rules:

$$\frac{\begin{matrix}[curry\,A] \\ A\end{matrix}}{curry\,A}\;curry\,\mathrm{I} \qquad \frac{curry\,A \qquad curry\,A}{A}\;curry\,\mathrm{E}$$

where $curry$ is a unary logical connective. To obtain the original paradoxical rules ρ^*I and ρ^*E, all we need to do is replace in these rules the arbitrary A with $\bot$ and $curry\,A$ with ρ^*.[18]

5 Conclusion

In this paper I tried to show that the analysis of the paradoxical proposition ρ^* utilizing constructive type theory (CTT) suggested by Tranchini (2019) is problematic because this proposition cannot be correctly formed in CTT, let alone proven. In other words, in CTT we are not able to properly justify the judgment $\rho^* : prop$. The same seems to apply to the other discussed variants of ρ^*, namely ρ, σ and τ.

The main issue with ρ^* lies in the circular nature of its introduction rule, more specifically, in the fact that there is a negative occurrence of ρ^* in its premise. This clashes with the general justification scheme of constructive type theory: formation rules, which tell us how to form new propositions, should be justified by the corresponding introduction rules, which tell us what these propositions mean, i.e., how to prove them. In the case of the proposition ρ^*, this justification requirement is, however, not met, since the introduction rule that should explain the meaning of ρ^* presupposes that we already understand it. Consequently, the formation rules cannot, strictly speaking, be understood as justified.

Therefore, we reached the conclusion that what is paradoxical about ρ^* from the perspective of CTT is rather whether or not it is a proposition at all, not that it can be proven and disproven at the same time.

Although some of the issues we have dealt with here might seem rather technical, from a more general point of view this paper is meant as a contribution to the general discussion concerning the nature of introduction rules. And even though at this time I am not yet ready to open the fundamental

[18] I thank one of the reviewers for this remark.

question "What is an introduction rule?", whatever the answer will be, it will have to take into account the matters discussed in this paper, i.e., how to deal with circularity, or more specifically, with negative occurrences within premises of introduction rules.

References

Bertot, Y., & Castéran, P. (2004). *Interactive Theorem Proving and Program Development*. Berlin: Springer.

Chlipala, A. (2013). *Certified Programming with Dependent Types*. Cambridge, MA: MIT Press.

Coquand, T. (1992). The paradox of trees in type theory. *BIT*, *32*(1), 10–14.

Curry, H. B., & Feys, R. (1958). *Combinatory Logic* (Vol. 1). Amsterdam: North-Holland Publishing Company.

Dummett, M. (1991). *The Logical Basis of Metaphysics*. London: Duckworth.

Dybjer, P. (1994). Inductive families. *Formal Aspects of Computing*, *6*(4), 440–465.

Dybjer, P. (1997). Representing inductively defined sets by wellorderings in Martin-Löf's type theory. *Theoretical Computer Science*, *175*, 329–335.

Dyckhoff, R. (2016). Some remarks on proof-theoretic semantics. In T. Piecha & P. Schroeder-Heister (Eds.), *Advances in Proof-Theoretic Semantics* (Vol. 43, pp. 79–93). Cham: Springer.

Frege, G. (1893). *Grundgesetze der Arithmetik I.* Jena: Hermann Pohle.

Gentzen, G. (1969). *The collected papers of Gerhard Gentzen*. Amsterdam: North-Holland Publishing Company.

Howard, W. A. (1980). The formulae-as-types notion of construction. In H. B. Curry, J. R. Hindley, & J. P. Seldin (Eds.), *To H. B. Curry: Essays on Combinatory Logic, Lambda Calculus, and Formalism.* London: Academic Press.

Klev, A. (2019a). A comparison of type theory with set theory. In S. Centrone, D. Kant, & D. Sarikaya (Eds.), *Reflections on the Foundations of Mathematics* (pp. 271–292). Cham: Springer.

Klev, A. (2019b). The justification of identity elimination in Martin-Löf's type theory. *Topoi*, *38*(3), 577–590.

Lorenzen, P. (1955). *Einführung in die operative Logik und Mathematik*. Berlin: Springer.

Martin-Löf, P. (1971). Hauptsatz for the intuitionistic theory of iterated inductive definitions. In J. E. Fenstad (Ed.), *Proceedings of the Second Scandinavian Logic Symposium (Oslo 1970)* (pp. 197–215). Amsterdam: North-Holland Publishing Company.

Martin-Löf, P. (1984). *Intuitionistic Type Theory*. Napoli: Bibliopolis.

Mendler, P. F. (1987). *Inductive definition in type theory* (Unpublished doctoral dissertation). Cornell University.

Pezlar, I. (2020). The placeholder view of assumptions and the Curry–Howard correspondence. *Synthese*.

Prawitz, D. (1965). *Natural Deduction: A Proof-theoretical Study*. New York: Dover Publications, Incorporated.

Prior, A. (1960). The runabout inference ticket. *Analysis*, *21*(1), 38–39.

Read, S. (2010). General-elimination harmony and the meaning of the logical constants. *Journal of Philosophical Logic*, *39*(5), 557–576.

Schroeder-Heister, P. (2012). Proof-theoretic semantics, self-contradiction, and the format of deductive reasoning. *Topoi*, *31*(1), 77–85.

Tennant, N. (1978). *Natural Logic*. Edinburgh: Edinburgh University Press.

Tennant, N. (1982). Proof and paradox. *Dialectica*, *36*(2–3), 265–296.

Tranchini, L. (2014). Proof-theoretic semantics, paradoxes and the distinction between sense and denotation. *Journal of Logic and Computation*, *26*(2), 495–512.

Tranchini, L. (2019). Proof, meaning and paradox: Some remarks. *Topoi*, *38*(3), 591–603.

Ivo Pezlar
Czech Academy of Sciences, Institute of Philosophy
The Czech Republic
E-mail: pezlar@flu.cas.cz