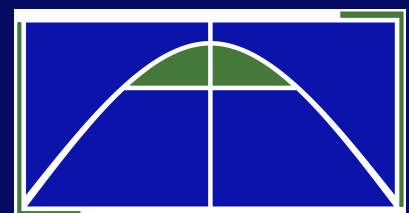


# **Seminário Lógica no Avião**

**2013 - 2018**

**Organizado por Edgar Almeida,  
Alexandre Costa-Leite e  
Rodrigo Freire**

**Lógica no Avião**



## **Conselho Editorial**

**FERNANDO FERREIRA**

Departamento de Matemática  
Universidade de Lisboa

**FRANCISCO MIRAGLIA**

Departamento de Matemática  
Universidade de São Paulo

**GRAHAM PRIEST**

Department of Philosophy  
The City University of New York

**JOHAN VAN BENTHEM**

Department of Philosophy Stanford University  
Tsinghua University  
University of Amsterdam

**MATTEO VIALE**

Dipartimento di Matematica “Giuseppe Peano”  
Università di Torino

Edgar Almeida, Alexandre Costa-Leite, Rodrigo Freire (Orgs.),  
Seminário Lógica no Avião: 2013 - 2018, Brasília: Lógica no  
Avião, 2019.

Série n, Volume 1

I.S.B.N. 978-65-00-06187-1

DOI: 10.21452/LnA\_serie\_n\_v01\_book\_seminario-logica-no-aviao-  
2013-2018\_195fl\_org.ea.rf.acl

Short DOI: <http://doi.org/c768>

Obra publicada com o apoio do PPGFIL/UnB.



## PREFÁCIO EDITORIAL

Este livro é constituído de contribuições originais de palestrantes do *Seminário Permanente Lógica no Avião* que visitaram Brasília entre o primeiro semestre de 2013 e o primeiro semestre de 2018. O seminário ocorre, desde o seu início, na Asa Norte, no Plano Piloto da cidade de Brasília, onde está situado o Campus Darcy Ribeiro da UnB. Uma interpretação possível da forma do Plano Piloto remete a um avião, o que inspirou o nome do seminário. Este contou com cinquenta palestras no período inicial de cinco anos, um volume de contribuições maior do que podemos colocar em um livro. Temos aqui, então, apenas onze contribuições divididas em duas partes representativas da atividade do seminário.

A primeira parte do livro contém artigos dedicados aos fundamentos da lógica e da matemática, que é a linha temática com maior atividade no seminário. Inicialmente, encontramos três investigações lógicas sobre teorias fundacionais - aritmética e teoria de conjuntos. Os artigos de Priest e de Martinez e Venturi exploram modelos para aritméticas e teorias de conjuntos paraconsistentes, enquanto o artigo de Ferreira é dedicado à clássica investigação lógica sobre os fundamentos da aritmética conduzida por Frege. Encontramos em seguida duas análises conceituais relacionadas à lógica intuicionista. O artigo de Haeusler e Rademaker analisa o tema das obrigações contrárias ao dever, e o artigo de Campos e Rodrigues analisa o princípio da explosão, ambos no contexto do intuicionismo. Fechando a primeira parte, temos os artigos de Mariano e de Souza dedicados a métodos de teoria de categorias no estudo dos sistemas lógicos.

Fundamentos da física e metafísica determinam a segunda linha temática principal do seminário e definem a segunda parte do livro. Abre esta parte o artigo de Krause acerca da interpretação da mecânica quântica. Em seguida, temos os artigos de Balashov e Boccardi em filosofia do tempo e do movimento. Fechando o volume, encontramos o ensaio de Durante sobre o conteúdo lógico das divergências metafísicas.

Agradecemos a todos os participantes do seminário, tanto aqueles que realizaram palestras quanto aqueles que estiveram presentes em tais atividades. Devemos, também, agradecimentos a Gustavo Schmidt, pela ajuda com a produção deste livro, e a Luiza Ramos, pela ajuda na organização de algumas palestras. Ainda, registramos nossa gratidão às pessoas que colaboraram indiretamente com o projeto. Sem a contribuição de todos os envolvidos não seria possível manter a regularidade que o seminário apresenta. A lista com as cinquenta primeiras palestras pode ser encontrada no final deste livro.

Edgar Almeida, Alexandre Costa-Leite e Rodrigo Freire.  
Brasília, junho de 2019.

# SUMÁRIO

## PARTE 1 - FUNDAMENTOS DA LÓGICA E DA MATEMÁTICA

### **Graham Priest**

Gödel's Theorem and Paraconsistency.....1

### **Santiago Jockwich Martinez & Giorgio Venturi**

Linear models and set theory.....9

### **Fernando Ferreira**

Hesitações na aritmética de Frege.....22

### **Edgar Campos & Abilio Rodrigues**

Some remarks on the validity of the principle of explosion in intuitionistic logic.....29

### **Edward Hermann Haeusler & Alexandre Rademaker**

On How Kelsenian Jurisprudence and Intuitionistic Logic help to avoid Contrary-to-Duty paradoxes in Legal Ontologies.....44

### **Hugo Luiz Mariano**

Alguns comentários sobre a questão da identidade entre lógicas.....60

### **Edelcio Gonçalves de Souza**

Categorias com morfismos verdade e prototopos.....73

## PARTE 2 - FUNDAMENTOS DA FÍSICA E METAFÍSICA

### **Décio Krause**

O gato de Schrödinger *não* está vivo e morto antes da medição.....104

### **Yuri Balashov**

The Common Present in a Block Universe.....123

### **Emiliano Boccardi**

Change and Contradiction: a criticism of the Hegelian account of motion.....135

### **Daniel Durante**

Nenhuma Divergência Metafísica Sem Incompatibilidade Lógica.....149

## PALESTRAS

Lista das palestras ocorridas entre 2013 e 2018.....188

PARTE 1

FUNDAMENTOS DA LÓGICA E DA MATEMÁTICA

# GÖDEL'S THEOREM AND PARACONSISTENCY

Graham Priest<sup>1</sup>

## 1. To Begin With

**1.1 Introduction.** Much has been written about Gödel's First Incompleteness Theorem. Nearly always, this is on the assumption that the logic of the theory of arithmetic in question is classical—or at least intuitionistic—logic. Rarely is the possibility that a paraconsistent logic is used explored. Yet the use of a paraconsistent logic puts a distinctively new spin on matters. The point of the present paper is to spell some of this out.

I shall not assume that people are familiar with the elements of paraconsistent logic—or at least, of the paraconsistent logic I will deploy. So before we get to paraconsistent arithmetic, I will spell this out. We will then be in a position to bring this to bear on Gödel's Theorem. Mostly, what I say will be of a technical nature; but I will permit myself a few philosophical comments towards the end.<sup>2</sup>

**1.2 Gödel's Theorem.** But first, some background.

If one looks at textbook statements of Gödel's Theorem, one will normally find something like this:

- Any axiomatic theory of arithmetic, with appropriate expressive capabilities, is incomplete—in the sense that there is something true (in the standard model of arithmetic) that cannot be proved.

This is inaccurate. What the proof of the theorem actually shows is this:

- Any axiomatic theory of arithmetic, with appropriate expressive capabilities, is *either* incomplete—in that sense—*or* inconsistent.

Of course, the inconsistent possibility is usually ignored, since in classical logic inconsistency implies triviality, and so this possibility is uninteresting. With a paraconsistent logic, it is, however.

It might be thought that the proof of the Theorem depends on classical logic. It does not. In fact, it makes virtually no assumptions about the underlying logic of the theory. Let me show this.

The theorem can be proved for any theory,  $T$ , which satisfies the following four conditions:

---

<sup>1</sup> Departments of Philosophy, the CUNY Graduate Center, and the University of Melbourne. E-mail: [priest.graham@gmail.com](mailto:priest.graham@gmail.com).

<sup>2</sup> This is a written up version of a talk that has been given in various places, including a meeting of the Swiss Graduate Society for Logic and the Philosophy of Science (Bern University), and the conferences Ouroboros 2018: Formal Criteria of Self-Reference in Mathematics and Philosophy (University of Bonn), and As Formas da Razão (Universidade de Brasília).

- [A] Gödel codes are assigned to syntactic entities, such as formulas and proofs. If  $n$  is a number, write its numeral as  $\mathbf{n}$ . If  $A$  is a formula with code  $n$ , write  $\langle A \rangle$  for  $\mathbf{n}$ .
- [B] There is a formula with two free variables,  $B(x, y)$ , which defines the proof relation of  $T$ . That is:
  - (i) if  $n$  is the code of a proof of  $A$  in  $T$  then  $B(\mathbf{n}, \langle A \rangle)$  is true in the standard model
  - (ii) if  $n$  is not the code of a proof of  $A$  in  $T$  then  $\neg B(\mathbf{n}, \langle A \rangle)$  is true in the standard model
- [C] Define  $Prov\ y$  as  $\exists x B(x, y)$ . Then  $Prov$  is a proof predicate for  $T$ . That is:
  - if  $T \vdash A$  then  $T \vdash Prov\langle A \rangle$
- [D] There is a formula,  $G$ , of the form  $\neg Prov\langle G \rangle$ .<sup>3</sup>

Given these assumptions, we can now establish Gödel's Theorem in the disjunctive form in which I stated it.

- By [D], if  $T \vdash G$  then  $T \vdash \neg Prov\langle G \rangle$ .
- By [C],  $T \vdash G$  then  $T \vdash Prov\langle G \rangle$ .
- Hence, if  $T \vdash G$ ,  $T$  is inconsistent.

Suppose that  $T$  is consistent.

- Then  $T \not\vdash G$ .
- That is, no number is the code of a proof of  $G$ .
- By [Bii], for any  $n$ ,  $\neg B(\mathbf{n}, \langle G \rangle)$  is true in the standard model.
- Hence,  $\forall x \neg B(x, \langle G \rangle)$  is true in the standard model.
  - As, then, is  $\neg \exists x B(x, \langle G \rangle)$ .
  - That is,  $\neg Prov\langle G \rangle$ .
  - That is,  $G$ .
- Hence,  $G$  is true in the standard model.

So if  $T$  is consistent, it is incomplete. Contrapositively, if  $T$  is complete, it is inconsistent.

Notice that virtually nothing is assumed in this argument about the logic of  $T$ .

## 2. Paraconsistent Logic

**2.1 Semantics.** Let us now turn to details concerning paraconsistent logic. The paraconsistent logic in question is  $LP$ . This may be set up in many ways. For our purposes, the following is most the appropriate.<sup>4</sup>

The language is a standard first-order language. We take  $A \supset B$  to be defined in the usual way, as  $\neg A \vee B$ . An interpretation for the language is a structure  $\mathfrak{A} = \langle D, \delta \rangle$  such that:

<sup>3</sup> Standard proofs of the theorem usually do not deliver quite this. What they deliver is a formula,  $G$ , such that  $G \equiv \neg Prov\langle G \rangle$  (and in a paraconsistent context,  $\equiv$  does not detach). However, what follows can be modified to take account of this fact with only minor changes. In any case, there are various ways to obtain a literal identity. My favourite is as follows. (I'm pretty sure that the idea is due to Saul Kripke, though I forget where I first learned it.) Extend the language of arithmetic by a new constant,  $\mathbf{k}$ . Assign Gödel codes to this language in the usual way. Let the code of  $\neg \exists x Prov\ \mathbf{k}$  be  $k$ . Now, consider an interpretation of this language which is the same as the standard interpretation, except that the denotation of  $\mathbf{k}$  is  $k$ . In this interpretation,  $\mathbf{k} = \langle \neg \exists x Prov\ \mathbf{k} \rangle$ .

<sup>4</sup> See Priest (2006a), ch. 5, and Priest (2008), chs. 7, 21.

- $D$  is a non-empty domain (of quantification).
- For every constant,  $c$ ,  $\delta(c) \in D$ .
- For every  $n$ -place function symbol,  $f_n$ ,  $\delta(f_n)$  is a function from  $D^n$  to  $D$ .
- For every  $n$ -place predicate,  $P_n$ ,  $\delta(P_n)$  is a pair,  $\langle \delta^+(P_n), \delta^-(P_n) \rangle$  such that:
  - $\delta^+(P_n) \cup \delta^-(P_n) = D^n$
- $\delta^+(=) = \{ \langle d, d \rangle : d \in D \}$

Let  $\mathfrak{A} \models^+ A$  and  $\mathfrak{A} \models^- A$  mean that  $A$  is true and false in  $\mathfrak{A}$ , respectively. Then, given any interpretation,  $\mathfrak{A}$ :

- $\delta(f_n t_1 \dots t_n) = \delta(f_n)(\delta(t_1), \dots, \delta(t_n))$
- $\mathfrak{A} \models^+ P_n t_1 \dots t_n$  iff  $\langle \delta(t_1), \dots, \delta(t_n) \rangle \in \delta^+(P_n)$
- $\mathfrak{A} \models^- P_n t_1 \dots t_n$  iff  $\langle \delta(t_1), \dots, \delta(t_n) \rangle \in \delta^-(P_n)$
- $\mathfrak{A} \models^+ \neg A$  iff  $\mathfrak{A} \models^- A$
- $\mathfrak{A} \models^- \neg A$  iff  $\mathfrak{A} \models^+ A$
- $\mathfrak{A} \models^+ A \wedge B$  iff  $\mathfrak{A} \models^+ A$  and  $\mathfrak{A} \models^+ B$
- $\mathfrak{A} \models^- A \wedge B$  iff  $\mathfrak{A} \models^- A$  or  $\mathfrak{A} \models^- B$
- $\mathfrak{A} \models^+ A \vee B$  iff  $\mathfrak{A} \models^+ A$  or  $\mathfrak{A} \models^+ B$
- $\mathfrak{A} \models^- A \vee B$  iff  $\mathfrak{A} \models^- A$  and  $\mathfrak{A} \models^- B$

To state the truth/falsity conditions for quantified statements, we assume that the language has been augmented with a new constant,  $k_d$ , for every  $d \in D$ , such that  $\delta(k_d) = d$ .<sup>5</sup> We then have:

- $\mathfrak{A} \models^+ \forall x A$  iff for all  $d \in D$   $\mathfrak{A} \models^+ A_x(k_d)$
- $\mathfrak{A} \models^- \forall x A$  iff for some  $d \in D$   $\mathfrak{A} \models^- A_x(k_d)$
- $\mathfrak{A} \models^+ \exists x A$  iff for some  $d \in D$   $\mathfrak{A} \models^+ A_x(k_d)$
- $\mathfrak{A} \models^- \exists x A$  iff for all  $d \in D$   $\mathfrak{A} \models^- A_x(k_d)$

For closed formulas,  $A$ , and sets of closed formulas,  $\Sigma$ :

- $\mathfrak{A}$  is a *model* of  $A$  iff  $\mathfrak{A} \models^+ A$
- $\mathfrak{A}$  is a *model* of  $\Sigma$  iff  $\mathfrak{A} \models^+ A$ , for every  $A \in \Sigma$
- $\Sigma \models_{LP} A$  iff every model of  $\Sigma$  is a model of  $A$

Note that all of this is *exactly* the same as classical logic ( $CL$ ), except that the semantics of  $LP$  allow truth and falsity to overlap.

Next, some simple properties of  $LP$ . Call an interpretation *classical* if  $\delta^+(P_n) \cap \delta^-(P_n) = \emptyset$ , for every  $P_n$ . Then clearly, every classical interpretation is an  $LP$  interpretation. Hence:

- $\Sigma \models_{LP} A \Rightarrow \Sigma \models_{CL} A$

However:

- $\{A, \neg A\} \not\models_{LP} B$

Hence:

- $\Sigma \models_{CL} A \not\Rightarrow \Sigma \models_{LP} A$

But:<sup>6</sup>

- $\emptyset \models_{CL} A \Leftrightarrow \emptyset \models_{LP} A$

<sup>5</sup> This is not essential. It is just a simple way of avoiding talk of satisfaction.

<sup>6</sup> For the proof, see Priest (2006a), ch. 5.



**2.2 Collapsed Models.** We now come to an important piece of *LP* metatheory: the Collapsing Lemma.

Let  $\mathfrak{A}$  be any *LP* interpretation (and so any classical interpretation), and let  $\sim$  be any equivalence relation on the domain of  $\mathfrak{A}$ , which is also a congruence relation on the interpretations of the function symbols in the language. That is:

- $d_1 \sim e_1, \dots, d_n \sim e_n \Rightarrow \delta(f_n)(d_1, \dots, d_n) \sim \delta(f_n)(e_1, \dots, e_n)$

If  $d \in D$ , let  $[d]$  be the equivalence class of  $d$ .

Now define the *collapsed interpretation*,  $\tilde{\mathfrak{A}} = \langle \tilde{D}, \tilde{\delta} \rangle$ , as follows:

- $\tilde{D} = \{[d] : d \in D\}$
- For every constant,  $c$ ,  $\tilde{\delta}(c) = [c]$
- For every  $n$ -place function symbol,  $f_n$ ,  $\tilde{\delta}(f_n)([d_1], \dots, [d_n]) = [\delta(f_n)(d_1, \dots, d_n)]$
- For every  $n$ -place predicate,  $P_n$ :
  - $\langle [d_1], \dots, [d_n] \rangle \in \tilde{\delta}^+(P_n)$  iff for some  $e_1 \sim d_1, \dots, e_n \sim d_n$   $\langle e_1, \dots, e_n \rangle \in \delta^+(P_n)$
  - $\langle [d_1], \dots, [d_n] \rangle \in \tilde{\delta}^-(P_n)$  iff for some  $e_1 \sim d_1, \dots, e_n \sim d_n$   $\langle e_1, \dots, e_n \rangle \in \delta^-(P_n)$

It is easy to check that the collapsed interpretation is indeed an *LP* interpretation. In effect, the collapse identifies all the members of an equivalence class, to deliver an object which has all the properties of the things in it.

One may now establish the Collapsing Lemma:

If  $\mathfrak{A}$  is any *LP* interpretation, and  $\tilde{\mathfrak{A}}$  is any collapse then:

- $\tilde{\delta}(t) = [\delta(t)]$
- If  $\mathfrak{A} \Vdash^+ A$  then  $\tilde{\mathfrak{A}} \Vdash^+ A$
- If  $\mathfrak{A} \Vdash^- A$  then  $\tilde{\mathfrak{A}} \Vdash^- A$

In other words, when an interpretation is collapsed, truth and falsity values are preserved (though other things may become true or false in the process). The proof of the Lemma is by the appropriate inductions, and I omit it here.<sup>7</sup>

### 3. Inconsistent Arithmetic

**3.1 Inconsistent Models.** We can now turn to Gödel's theorem. To do so, we need to apply the Collapsing Lemma to establish some facts about models of inconsistent arithmetic.<sup>8</sup>

In what follows, let the language be the standard one for first-order arithmetic (with one constant, 0, and function symbols for successor, addition, and multiplication). Let  $\mathbf{N}$  be the set of sentences in this language true in the standard (classical) model of arithmetic. Let  $\mathfrak{M}$  be any model of  $\mathbf{N}$ .

If  $\tilde{\mathfrak{M}}$  is any collapsed interpretation,  $\tilde{\mathfrak{M}}$  is a model of  $\mathbf{N}$ , by the Collapsing Lemma. Let me give a couple of examples of such collapsed models.

*Example 1*

- $\mathfrak{M}$  is any classical non-standard model of arithmetic

<sup>7</sup> Details can be found in Priest (2006a), p. 229 f.

<sup>8</sup> For this, and much further discussion, see Priest (2006a), ch. 17, and the references cited therein.

- $d \sim e$  iff ( $d$  and  $e$  are standard and  $d = e$ ) or ( $d$  and  $e$  are non-standard)

What  $\sim$  does, in effect, is to leave all the standard numbers alone, but identify all the non-standard numbers. It is easy to check that  $\sim$  is an equivalence relation, and a congruence with respect to the function symbols. We may draw the diagram of the successor relation in the collapsed model as follows:

$$0 \rightarrow 1 \rightarrow 2 \rightarrow \dots \rightarrow i \quad \curvearrowright$$

As is clear,  $\tilde{\mathfrak{M}} \models \exists x x = x'$ , but  $\tilde{\mathfrak{M}} \not\models 0 = 0'$ . Hence,  $\tilde{\mathfrak{M}}$  is a model of  $\mathbf{N}$ , and the set of things true in it is inconsistent but non-trivial (i.e., not the set of all formulas).

*Example 2*

- $\mathfrak{M}$  is the standard model of arithmetic
- $n$  and  $p$  are a natural numbers greater than 1.
- $d \sim e$  iff ( $d, e < n$  and  $d = e$ ) or ( $d, e \geq n$  and  $d = e \pmod{p}$ )

This relation leaves numbers less than  $n$  alone, but numbers greater than or equal to  $n$  behave as in arithmetic mod  $p$ . Again, is is easy to check that this is an equivalence relation, and a congruence relation on the function symbols. We may depict the collapsed interpretation as follows.

$$\begin{array}{ccccccc} & & & & n+p-1 & \leftarrow & \dots & \leftarrow & n+3 \\ & & & & \downarrow & & & & \uparrow \\ 0 & \rightarrow & 1 & \rightarrow & \dots & \rightarrow & n & \rightarrow & n+1 & \rightarrow & n+2 \end{array}$$

As is clear,  $\tilde{\mathfrak{M}} \models \exists x x = x' \dots'$  (with  $p$  occurrences of the successor function symbol), and  $\tilde{\mathfrak{M}} \not\models 0 = 0'$ . Hence, again,  $\tilde{\mathfrak{M}}$  is a model of an inconsistent but non-trivial set of sentences that contains all the sentences true in the standard model. Moreover, in this case,  $\tilde{\mathfrak{M}}$  is finite. Hence, the set of sentences true in  $\tilde{\mathfrak{M}}$  is decidable. (The particular and universal quantifiers behave like finite disjunctions and conjunctions, respectively.) Hence, it is axiomatizable.

We see, then, that there are non-trivial but inconsistent axiomatic *LP* theories which contain all the truths in the standard model. There are, therefore axiomatic *LP* theories which are complete but inconsistent.<sup>9</sup> Thus, in the disjunctive statement of Gödel's Theorem stated in 1.2, one can no longer ignore the inconsistency disjunct.

**3.2 Back to the Gödel Sentence.** Let us now turn to look at the behaviour of the “Gödel sentence” in these inconsistent theories.

Let  $T$  be any axiomatic arithmetic such that  $T \supseteq \mathbf{N}$ . Let  $Prov x$  be the arithmetic predicate which defines  $T$  in the standard model. That is:

- If  $T \vdash A$  then  $Prov \langle A \rangle \in \mathbf{N}$ , so  $T \vdash Prov \langle A \rangle$
- If  $T \not\vdash A$  then  $\neg Prov \langle A \rangle \in \mathbf{N}$ , so  $T \vdash \neg Prov \langle A \rangle$

<sup>9</sup> I note that in the case of Example 2, the model is finite. But this is unnecessary. There are infinite models of the same kind, i.e., whose theories are axiomatic. This is proved in Paris and Sirokofskich (2008).

Let  $G$  be  $\neg Prov\langle G\rangle$ . Then either  $T \vdash G$  or  $T \not\vdash G$ . In either case,  $T \vdash \neg Prov\langle G\rangle$ . So  $T \vdash G$ , and  $T \vdash Prov\langle G\rangle$ . That is, the Gödel sentence and its negation are both provable—as opposed to the consistent case, where neither is provable.<sup>10</sup>

**3.3 Some Philosophical Issues.** So much for purely technical matters. Let me now make a few philosophical observations.

Behind Gödel’s proof of the Incompleteness Theorem—and of course, there are others—there is a recognisable paradox concerning the sentence ‘this sentence is not provable’, which is clearly in the same family as the Liar. We may call this *Gödel’s Paradox*, and it goes as follows. Let  $G$  be  $\neg Prov\langle A\rangle$ .

- Suppose that  $Prov\langle G\rangle$
- Then  $G$  (what is provable is true)
- That is,  $\neg Prov\langle G\rangle$
- Hence,  $\neg Prov\langle G\rangle$  (is proved)
- That is,  $G$
- So  $Prov\langle G\rangle$  (we have just proved it)

Now, of course, this argument cannot be formalised in the arithmetic if it is consistent: the second line cannot be enforced, however natural it is. However, in a dialethic solution to the Liar Paradox, the thought is that there is nothing wrong with the reasoning which produces contradiction. It is just fine, and we should accept the contradiction it delivers (though the use of a paraconsistent logic prevents an arbitrary conclusion following). If this is correct then, it would seem, we should accept the same conclusion in the case of the Gödel’s Paradox. The inconsistent arithmetic, then, gets things exactly right.

If one asks why one cannot push through the paradoxical argument in the case of consistent arithmetics, the answer is that Löb’s Theorem gets in the way. For the theorem tells us that, for any  $A$ , if  $\vdash Prov\langle A\rangle \supset A$  then  $\vdash A$ . Hence, we cannot infer  $A$  from  $Prov\langle A\rangle$ , for arbitrary  $A$ .

Now, in truth, there is already something strange about Löb’s Theorem. Given that the axiom system is sound, of course one ought to be able to infer  $A$  from the fact that  $A$  is proved. Indeed, for any  $A$ ,  $Prov\langle A\rangle \supset A$  is true in the standard model!<sup>11</sup> Intuitively, one should expect all the Löb sentences to hold in the correct arithmetic. And they do in the inconsistent arithmetics. To see this, note that for any  $A$ , either  $T \vdash A$  or  $T \not\vdash A$ . In the first case, obviously  $T \vdash Prov\langle A\rangle \supset A$ . In the second case,  $T \vdash \neg Prov\langle A\rangle$ , so  $T \vdash Prov\langle A\rangle \supset A$ .

<sup>10</sup> It might be wondered what happens to Gödel’s Second Theorem in this context. This theorem states that given a consistent axiomatic arithmetic of the kind we are dealing with, the canonical statement of its consistency cannot be proved in the theory. In the case of inconsistent arithmetics, since they are inconsistent, one would not expect the theory to be able to prove its consistency. However, if the theory is non-trivial, one might hope that the theory can prove its own non-triviality. Indeed in the case of the inconsistent axiomatic arithmetics we are considering, it can. For take some unprovable sentence,  $A$ . Then since  $T \not\vdash A$ ,  $\neg Prov\langle A\rangle \in \mathbf{N}$ , and so  $T \vdash \neg Prov\langle A\rangle$ . (Warning: in some theories, we might be able to prove  $Prov\langle A\rangle$  as well. See, further, Priest (2006a), 17.4.)

<sup>11</sup> In case anyone is puzzled by this technically, note that one can add all instances of the Löb sentence to obtain a consistent theory. The cost is that  $Prov$  no longer represents provability in this new system.

#### 4. Objections

Let me end with a brief mention of a few philosophical objections one might be tempted to make at this point.

Perhaps the first thought would be that one cannot prove enough in the inconsistent arithmetics. After all, *modus ponens* for  $\supset$  is not valid. That thought can be set aside quickly, though. Everything true in the standard model is provable in the theory: everything one could want.

But the next obvious thought is that one can prove too much. True, we have everything true in the standard model; but we have some of their negations too. Why, however, should such contradictions be a problem? One reason one might give is that, by the Principle of Non-Contradiction, contradictions are not true, so one should not want them to be provable. Clearly, the objection turns on the correctness of the Principle of Non-Contradiction, and so now the debate must turn to the question of what reason one has for supposing this to be true. This is not the place to go into the matter. Let me just say that arguments for the unrestricted validity of the Principle are exceptionally hard to find.<sup>12</sup>

The next thought is a more subtle one. We apply arithmetic in science for many purposes. Contradictions in arithmetic would ruin these applications. This does not follow, however. Thus, suppose that the inconsistent arithmetic is given as in Example 2. Then all the equations for numbers less  $n$  behave consistently.<sup>13</sup> So any sentence where the quantifiers are bounded by a number less than  $n$  also behave consistently. Now, for practical purposes, we are rarely concerned with *really* large numbers. Thus suppose that there is an inconsistent natural number, and so a least,  $m + 1$ . Suppose that this is so large that it has no empirical meaning. (Maybe  $m$  is  $A(n, n)$ , where  $A$  is the Ackermann function, and  $n$  is the number of elementary particles in the cosmos.) Any sentence with quantifiers bounded by  $m$  is consistent. So the contradictions have no practical implications.<sup>14</sup>

#### 5. Conclusion

Whatever one makes of these matters, it is clear that applying a paraconsistent logic to arithmetic opens up a whole new dimension on matters concerning Gödel's Incompleteness Theorem. Indeed, allowing for the possibility of inconsistency-tolerant theories, does this in many areas. But Gödel's Theorem is undoubtedly one of the most profound pieces of mathematics in the last 100 years. The new dimensions concerning this are, then, equally profound.

---

<sup>12</sup> For just the start of the debate, see Priest (1989). See also the essays in Priest (200b).

<sup>13</sup> And of course, in Example 1, all the natural numbers behave consistently.

<sup>14</sup> For some further discussion, see Priest (1994).

## Bibliography

- [1] Paris, J. B., and Sirokofskich, A. (2008), ‘On *LP* Models of Arithmetic’, *Journal of Symbolic Logic* 73: 212-226.
- [2] Priest, G. (1994), ‘What could the Least Inconsistent Number be?’, *Logique et Analyse* 37: 3-12.
- [3] Priest, G. (1998), ‘What’s so Bad about Contradictions?’, *Journal of Philosophy* 95: 410–26; reprinted as ch. 1 of Priest, G., Beall, J. C., and Armour-Garb, B. (eds.), *New Essays on the Law of Non-Contradiction*, Oxford: Oxford University Press, 2004.
- [4] Priest, G. (2006a), *In Contradiction*, 2nd edn, Oxford: Oxford University Press.
- [5] Priest, (2006b), *Doubt Truth to be a Liar*, Oxford: Oxford University Press.
- [6] Priest, G. (2008), *Introduction to Non-Classical Logic*, 2nd edn, Cambridge: Cambridge University Press.

# LINEAR MODELS AND SET THEORY

Santiago Jockwich Martinez<sup>1</sup> & Giorgio Venturi<sup>2</sup>

## Abstract

In this paper we review the approach to algebra-valued models and we introduce an infinite class of natural models for paraconsistent set theory. We then present several paraconsistent set theories based on Logics of Formal Inconsistency (LFI), known in the literature as LFI-set theories, and we show that there is no natural model that validates these systems. We therefore suggest to abandon these theories.

## 1. Introduction

There are two possible approaches to paraconsistent set theory. The first one, that we may call *Cantorian*, stems from the recognition that the troubles caused by the foundational paradoxes is not ascribable to the basic principles of the theory but to the underlying logic. This approach calls for a revision of the logic and proposes new logical tools to escape the triviality that originates from set-theoretical antinomies. Generally, this kind of approach aims at formalizing set theory by means of two simple principles: Extensionality and unrestricted Comprehension. Examples of these theories can be found in [3], [14], and [15]. While the proponents of these paraconsistent set theories consider their systems to be closer in spirit to Cantor's theory—and to the naïve idea that to every well-defined property corresponds a set—however their justification suffer of the same weakness that an intuitionist axiomatic system suffers with respect to Brouwer's intuitionism: how to consider appropriate the use of axioms and deductive rules for regimenting theories that in the intentions of their creators tended to escape formalization? In other words, how to conciliate the axiomatic method and a naïve approach to Cantor's theory of transfinite numbers?

In direct opposition to the Cantorian approach we may find what we may called *Zermelian* paraconsistent set theories. These are the axiomatic systems that, using an underlying paraconsistent logic, accept as a fundamental step in the development of set theory the axiomatization that Zermelo proposed in the 1908 paper *Investigations in the foundations of set theory I*. From this perspective, classical ZFC captures only partially the universe of set theory, while a paraconsistent approach is justified in that it offers a broader view of a theory which includes consistent and inconsistent totalities—to use a Cantorian phrasing. To put it more emphatically: there are more things in heaven and earth than are dreamt of in ZFC. Consequently, a Zermelian approach proposes axiomatic systems as close as possible to classical ZFC, but where the logical connective receive, instead, a paraconsistent interpretation. Example of these theories may be found in [8], [9], and [5]. Although both approaches have their pros and cons, we consider the Zermelian approach closer to set-theoretical practice.

It is undeniable that, from 1908 on, the axiomatization of set theory brought a clearer picture of the universe of all sets—after decades of works and the efforts of mathematicians

---

<sup>1</sup> University of Campinas, UNICAMP. Email: [santijoxi@hotmail.com](mailto:santijoxi@hotmail.com).

<sup>2</sup> University of Campinas, UNICAMP. Email: [gio.venturi@gmail.com](mailto:gio.venturi@gmail.com).

of the caliber of Zermelo, Fraenkel, Skolem, Von Neumann, and Gödel—in terms of the cumulative structure of  $\mathbf{V}$ . Moreover, on the philosophical side, Boolos [2] argued for a deep conceptual connection between a cumulative hierarchy structure, the iterative conception of set, and ZFC. For this reason, we believe that the Zermelian approach should not only be restricted to a system of axioms, but should also consider its semantic interpretation, in terms of a suitable cumulative hierarchy. This idea has been made concrete in [9], using *algebra-valued models*; i.e. a generalization of Boolean-valued models to structures whose underlying algebra is not necessarily Boolean. In [9], the authors introduced an algebra-valued model  $\mathbf{V}^{(\mathbf{PS}_3, *)}$ , based on the paraconsistent algebra  $(\mathbf{PS}_3, *)$ , first introduced in [10]. The approach of [9] has been recently extended in [12] to produce an infinite class of models, called *linear*, which include the model  $\mathbf{V}^{(\mathbf{PS}_3, *)}$  as a special case.

In this paper we investigate whether linear models are suitable structures for the set theories developed in [5]. The reason for asking this question is twofold. First, the linear models are the only known models for paraconsistent Zermelian set theories. Moreover, being the result of algebra-valued constructions, linear models can be seen as natural models for the Zermelian approach, because of their cumulative hierarchy structure. Second, and mostly, the algebra  $(\mathbf{PS}_3, *)$  is a Logic of Formal Inconsistency, as shown in [7]. Therefore it is natural to interpret the axioms of these axiomatic systems in a model like  $\mathbf{V}^{(\mathbf{PS}_3, *)}$ , which is based on the same logical background.

In the next pages we will show that the set theories developed in [5] are not valid in any linear model. This will therefore motivate the rejection of these theories. More precisely we will show that there is a deep conceptual issue in the formulation of these axiomatic systems, since they fall short of capturing inconsistent totalities. In short, these systems do not comply with the task for which they have been created; either they get a harmless interpretation of paraconsistency (i.e. classical ZF), or they are just false in the natural models of Zermelian paraconsistent set theories.

The article is structured as follows. In §2 we introduce the Boolean-valued technique for constructing models of set theory. In §3 we introduce the linear models for paraconsistent set theory. In §4 we introduce a family of Zermelian paraconsistent set theories based on the language of LFIs—as presented in [5]—and we show that linear models do not validate them. We conclude offering reasons to abandon these set theories.

## 2. Boolean valued models

The Boolean-valued model technique represents an alternative, algebraic, presentation of the forcing method. Forcing was invented by Paul Cohen [6] and is now considered the cornerstone of contemporary set theory. It is a tool for producing relative consistency results, that has been used to prove the independence of CH from the axioms of ZFC.

To recall the main ideas and to fix notation we briefly introduce the Boolean-valued model construction of a model of set theory.<sup>3</sup> Given a complete Boolean algebra  $\mathbb{B}$ , a Boolean-valued structure is defined by transfinite recursion:

$$\mathbf{V}_\alpha^{\mathbb{B}} = \{x \mid x \text{ is a function and } \text{ran}(x) \subseteq \mathbb{B} \text{ and there is } \zeta < \alpha \text{ with } \text{dom}(x) \subseteq \mathbf{V}_\zeta^{\mathbb{B}}\} \text{ and}$$

$$\mathbf{V}^{\mathbb{B}} = \{x \mid \exists \alpha (x \in \mathbf{V}_\alpha^{\mathbb{B}})\}.$$

<sup>3</sup> For reasons of space, we do not define notions that are standard in the literature. For a broader and more complete presentation, we defer the interested reader to [1].

Extending the language of set theory with constants for every element of  $\mathbf{V}^{\mathbb{B}}$ , we can define a new language  $\mathcal{L}^{\mathbb{B}}$ , as common in model theory, that allows one to talk about the Boolean valued structure  $\mathbf{V}^{\mathbb{B}}$ , by means of what are called  $\mathbb{B}$ -sentences. We can therefore generalize the notion of satisfaction. Indeed, it is possible to define a  $\mathbb{B}$ -evaluation function that assigns to every  $\mathbb{B}$ -sentence  $\sigma$ , its Boolean truth value  $\llbracket \sigma \rrbracket \in \mathbb{B}$ .<sup>4</sup> Hence, whenever  $\llbracket \varphi \rrbracket = 1_B$  (that is, the maximal element of the complete Boolean algebra  $\mathbb{B}$ ), then we say that  $\varphi$  is valid in  $\mathbf{V}^{\mathbb{B}}$ . The definition the  $\mathbb{B}$ -evaluation function not only resembles closely that of a truth definition, but, similarly, it cannot be fully defined within ZFC, because, as in the case of the truth predicate, the collection of all ordered pairs  $(\sigma, \llbracket \sigma \rrbracket^{\mathbb{B}})$  is not a definable class in ZFC.<sup>5</sup>

The Boolean-valued structures so constructed are non-standard in a very strong sense. Indeed, the interpretations of the basic notions of equality and membership have an ad hoc non-standard character. Given a complete Boolean algebra  $\mathbb{B}$  and  $u, v \in \mathbf{V}^{\mathbb{B}}$ ,

$$\llbracket u \in v \rrbracket = \bigvee_{y \in \text{dom}(v)} (v(y) \wedge \llbracket u = y \rrbracket)$$

and

$$\llbracket u = v \rrbracket = \bigwedge_{x \in \text{dom}(u)} (u(x) \Rightarrow \llbracket x \in v \rrbracket) \wedge \bigwedge_{y \in \text{dom}(v)} (v(y) \Rightarrow \llbracket y \in u \rrbracket).$$

Notice that the above evaluations are inter-defined by a simultaneous recursion and in terms of the algebraic structure of the Boolean algebra. On the other hand, the other logical connective are interpreted in the standard way.

$$\begin{aligned} \llbracket \perp \rrbracket^{\mathbb{B}} &= 0_B, \\ \llbracket \varphi \wedge \psi \rrbracket^{\mathbb{B}} &= \llbracket \varphi \rrbracket^{\mathbb{B}} \wedge_{\mathbb{B}} \llbracket \psi \rrbracket^{\mathbb{B}}, \\ \llbracket \varphi \vee \psi \rrbracket^{\mathbb{B}} &= \llbracket \varphi \rrbracket^{\mathbb{B}} \vee_{\mathbb{B}} \llbracket \psi \rrbracket^{\mathbb{B}}, \\ \llbracket \varphi \rightarrow \psi \rrbracket^{\mathbb{B}} &= \llbracket \varphi \rrbracket^{\mathbb{B}} \Rightarrow_{\mathbb{B}} \llbracket \psi \rrbracket^{\mathbb{B}}, \\ \llbracket \exists x \varphi(x) \rrbracket^{\mathbb{B}} &= \bigvee_{u \in \mathbf{V}^{\mathbb{B}}} \llbracket \varphi(u) \rrbracket^{\mathbb{B}}, \text{ and} \\ \llbracket \forall x \varphi(x) \rrbracket^{\mathbb{B}} &= \bigwedge_{u \in \mathbf{V}^{\mathbb{B}}} \llbracket \varphi(u) \rrbracket^{\mathbb{B}}, \end{aligned}$$

where on the left side of the equality are displayed logical connectives, while on the right side the operations of the algebra  $\mathbb{B}$ , and the negation is defined in terms of implication and  $\perp$ .

The interest of this construction lays in the fact that, due to the properties of Boolean algebras, all axioms of ZFC are valid in  $\mathbf{V}^{\mathbb{B}}$ , that is, if  $\sigma$  is an axiom of ZFC, then  $\llbracket \sigma \rrbracket^{\mathbb{B}} = 1_B$ . On the other hand, if  $\llbracket \sigma \rrbracket^{\mathbb{B}} \neq 0_B$ , we can say that  $\sigma$  is consistent with ZFC. Moreover, the versatility of this method consists in the possibility to obtain different consistency results just by changing the underlying algebra and thus constructing different models.

<sup>4</sup> Remember that  $0_B$  and  $1_B$ , respectively the smallest and largest elements of a Boolean algebra  $\mathbb{B}$ , may be interpreted as falsehood and truth; hence, all other elements of the algebra have the intuitive meaning of intermediate truth values.

<sup>5</sup> However, notice that, due to the definability in the ground model of the forcing relation, we have that, fixing a formula  $\varphi$ , the collection  $\{(\tau, \varphi(\tau)) \mid \tau \in \mathbf{V}^{\mathbb{B}}\}$  is a class definable in  $\mathbf{V}^{\mathbb{B}}$ .



### 3. Models for Zermelian set theories

DEFINITION 1. We call a poset  $(A; \leq)$  a meet semilattice if every two elements  $x, y \in A$  have an infimum, denoted by  $x \wedge y$ . If there also exists a supremum,  $x \vee y$ , for any two elements  $x, y \in A$ , then  $(A; \leq)$  is a lattice. We say that  $(A; \leq)$  is a bounded lattice if it is a lattice that has a greatest element  $1_A$  and a least element  $0_A$ . A lattice  $(A; \leq)$  is complete if the supremum  $\bigvee X$  and the infimum  $\bigwedge X$  exist for every  $X \subseteq A$ . A lattice is called distributive if it satisfies the distributivity law, that is,  $x \wedge (y \vee z) = (x \wedge y) \vee (x \wedge z)$  for all  $x, y, z$  in its universe.

In what follows, by an *algebra* we mean a complete bounded distributive lattice. A first step in the construction of natural models for Zermelian paraconsistent set theories has been undertaken in [9], modifying the underlying algebra of a Boolean-valued model construction. The authors of [9] individuated a specific class of algebra suited for this purpose: the reasonable implicative algebras (ri-algebras henceforth). Following [9], an algebra, in order to be a ri-algebra needs to have an operation  $\Rightarrow$  that satisfies the following properties:<sup>6</sup>

$$\mathbf{P1} \quad (x \wedge y) \leq z \text{ implies } x \leq (y \Rightarrow z)$$

$$\mathbf{P2} \quad y \leq z \text{ implies } (x \Rightarrow y) \leq (x \Rightarrow z)$$

$$\mathbf{P3} \quad y \leq z \text{ implies } (z \Rightarrow x) \leq (y \Rightarrow x)$$

$$\mathbf{P4} \quad ((x \wedge y) \Rightarrow z) = (x \Rightarrow (y \Rightarrow z))$$

But a further condition (or better a schema of conditions for every  $\varphi$  a negation free formula) is then required, in order validate all axioms of ZF.

$$\llbracket \forall x \varphi(x) \rrbracket^{\mathbb{A}} = \bigwedge_{x \in \text{dom}(u)} (u(x) \Rightarrow \llbracket \varphi(x) \rrbracket^{\mathbb{A}}) \quad (\mathcal{BQ}_\varphi)$$

THEOREM 2. ([9]) Any ri-algebra  $\mathbb{A}$  that satisfies  $\mathcal{BQ}_\varphi$ , for any negation free formula  $\varphi$ , is such that the model  $\mathbf{V}^{\mathbb{A}}$  validates the negation free fragment of ZF.

In [9] we find an example of an algebra, called  $\mathbf{PS}_3$ , for which the hypotheses—and thus the conclusion—of Theorem 2 hold. Thus,  $\mathbf{V}^{\mathbf{PS}_3}$  validates all axioms of ZF, with the axiom schemata restricted to negation free formulas.

$\Rightarrow$	1	$\frac{1}{2}$	0	$\vee$	1	$\frac{1}{2}$	0	$\wedge$	1	$\frac{1}{2}$	0
1	1	1	0	1	1	1	1	1	1	$\frac{1}{2}$	0
$\frac{1}{2}$	1	1	0	$\frac{1}{2}$	1	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	0
0	1	1	1	0	1	$\frac{1}{2}$	0	0	0	0	0

Figure 1. Connectives for  $\mathbf{PS}_3$

<sup>6</sup> A *caveat* for the reader: by an ri-algebra we mean what in [9] was called reasonable deductive implication algebra. We slightly changed the terminology for notational and expository convenience. We hope not to confuse the well-informed reader.

The main result of [9] consisted in showing that it is also possible to add a paraconsistent negation  $*$ , defined by  $1^* = 0$ ,  $\frac{1}{2}^* = \frac{1}{2}$  and  $0^* = 1$ , in such a way that  $\mathbf{V}^{(\text{PS}_3,*)}$  is a paraconsistent model of the negation free fragment of ZF (which will be abbreviated as NFF-ZF). More recently, in [12], this approach has been extended to show that  $\mathbf{V}^{(\text{PS}_3,*)}$  is only one element of a uniformly defined class of models for paraconsistent NFF-ZF.

DEFINITION 3. By  $\mathfrak{V}$  we indicate the class of linear algebras of the form  $\mathbb{L} = (L, \wedge, \vee, \neg, D, 0, 1)$ , where  $\neg x = \max\{y : x \wedge y = 0\}$  exists for every  $x \in L$  as well as  $Dx = \min\{y \in L : x \vee y = 1\}$  exists for every  $x \in L$ , such that the base set is order-isomorphic to an ordinal. By an  $\mathfrak{V}$ -algebra we mean a member of  $\mathfrak{V}$ .

THEOREM 4. ([12]) Given an  $\mathfrak{V}$ -algebra  $\mathbb{L}$ , it is possible to define an implication that not only makes it an ri-algebras, but that also satisfies  $\mathcal{BQ}_\varphi$ , for any negation free formula  $\varphi$ . Moreover, it is possible to define a paraconsistent negation in such a way that  $\mathbf{V}^{\mathbb{L}}$  is a model of paraconsistent ZF.

We refer to an  $\mathfrak{V}$ -algebra also as  $\mathbb{L} = (L; \wedge, \vee, \rightsquigarrow, *)$ , displaying the two operations  $\rightsquigarrow$  and  $*$  of Theorem 4. These operations are used to interpret the implication and the negation from the language of set theory. For this reason we call these algebraic operations, respectively, implication and negation.

LEMMA 5. ([12]) Let  $\mathbb{L}$  be a linear algebra of the form  $\mathbb{L} = (L; \wedge, \vee, \rightsquigarrow, *)$ . When  $|\mathbb{L}| = 2$  we get that  $\mathbf{V}^{\mathbb{L}} = \mathbf{V}^{\mathbb{B}_2}$ , where  $\mathbb{B}_2$  is the two-elements Boolean algebra, while when  $|\mathbb{L}| = 3$  we get that  $\mathbf{V}^{\mathbb{L}} = \mathbf{V}^{(\text{PS}_3,*)}$ .

The names negation and implication for the characteristic operations of linear algebras are well justified by the the next two lemmas that show their proximity to the classical case.

LEMMA 6. ([12]) The negation  $*$  of a linear algebra  $\mathbb{L}$  maps each intermediate value of the base set to itself, that is for all  $a \in L$  such that  $0 \neq a \neq 1$  we have  $a^* = a$ . Besides that  $0^* = 1$  and  $1^* = 0$ .

LEMMA 7. ([12]) For any  $\mathbb{L} \in \mathfrak{V}$  and for any  $x, y \in \mathbf{V}^{\mathbb{L}}$  the following holds:  $\llbracket \varphi(x) \rightsquigarrow \psi(y) \rrbracket = 0$  if and only if  $\llbracket \varphi(x) \rrbracket \neq 0$  and  $\llbracket \psi(y) \rrbracket = 0$ .  $\square$

#### 4. Logics of Formal Inconsistency and their set theories

In this section we briefly introduce the Zermelian set theories we will discuss in this paper and the logics on which they are based.

**4.1 Logics of Formal Inconsistency.** The LFIs form a family of paraconsistent logics that extend the syntactic approach of Da Costa to paraconsistency, by expressing consistency at the meta-linguistic level by means of a unary operator which extends the logical language. These logics were first introduced by João Marcos in his doctoral thesis [11].<sup>7</sup> More syntactically, we can define these logics as those which satisfy the following two conditions:

- (1) There is a set of formulas  $\Gamma$  and formulas  $\alpha$  and  $\beta$  such that  $\Gamma \cup \{\alpha, \neg\alpha\} \not\vdash \beta$ .

---

<sup>7</sup> For an up to date presentation of LFI see [4].

(2) Given a formula  $\alpha$ , there is a set of formulas  $\circ(\alpha)$ , uniquely determined by  $\alpha$ , such that for any set of formulas  $\Gamma$  and for any formula  $\beta$ ,  $\Gamma \cup \circ(\alpha) \cup \{\alpha, \neg\alpha\} \vdash \beta$ .

The principle (1) implies that LFIs are non-explosive logics, in the sense that a contradiction does not always entail arbitrary statements, i.e. the formula  $\alpha \rightarrow (\neg\alpha \rightarrow \beta)$  is not valid. On the other hand, the intuitive motivation of (2) is that negation is only explosive with respect to consistent formulas; therefore a theory that admits  $\circ(\alpha), \alpha, \neg\alpha$  is therefore trivial. This allows us to separate the sentences that make a theory explosive, from those that do not.

The *weakest* LFI is called *mbC* and consists of positive classical propositional logic ( $CPL^+$ ), plus a paraconsistent negation  $\neg$  and a consistency operator  $\circ$  which are added, together with a minimal set of rules to make it an LFI. In this case  $\circ(\alpha)$  is just the singleton  $\{\circ\alpha\}$ . Notice that the classical negation can be defined within *mbc* as  $\sim\alpha =_{df} \alpha \rightarrow (\circ\alpha \wedge \alpha \wedge \neg\alpha)$ . Furthermore, *mbc* constitutes a conservative extension of  $CPL^+$ , since every theorem of  $CPL^+$  can be recovered in *mbC*.

DEFINITION 8. *The logic mbC is defined over the language the classical propositional language extended by the  $\circ$ -operator and is axiomatized by the following Hilbert-style axioms:*

**Axiom schemas:**

$\alpha \rightarrow (\beta \rightarrow \alpha)$	<b>(Ax1)</b>
$(\alpha \rightarrow (\beta \rightarrow \gamma)) \rightarrow ((\alpha \rightarrow \beta) \rightarrow (\alpha \rightarrow \gamma))$	<b>(Ax2)</b>
$\alpha \rightarrow (\beta \rightarrow (\alpha \wedge \beta))$	<b>(Ax3)</b>
$(\alpha \wedge \beta) \rightarrow \alpha$	<b>(Ax4)</b>
$(\alpha \wedge \beta) \rightarrow \beta$	<b>(Ax5)</b>
$\alpha \rightarrow (\alpha \vee \beta)$	<b>(Ax6)</b>
$\beta \rightarrow (\alpha \vee \beta)$	<b>(Ax7)</b>
$(\alpha \rightarrow \gamma) \rightarrow ((\beta \rightarrow \gamma) \rightarrow ((\alpha \vee \beta) \rightarrow \gamma))$	<b>(Ax8)</b>
$\alpha \vee \neg\alpha$	<b>(Ax9)</b>
$\circ\alpha \rightarrow (\alpha \rightarrow (\neg\alpha \rightarrow \beta))$	<b>(Ax10)</b>

**Inference rule:**

$$\frac{\alpha \quad \alpha \rightarrow \beta}{\beta} \text{ (MP)}$$

Observe that **(Ax1 – Ax9)** plus **(MP)** constitute a Hilbert calculus over the signature for  $CPL^+$ , whereas **(Ax10)** is called the *gentle* law of explosion and characterizes primarily *mbC* as an **LFI**.

Next we will present the first order (with equality) version of *mbc*, the LFI  $Qmbc_{\approx}$ .

DEFINITION 9. *Let  $\Theta$  be a first-order signature with a  $\approx$ -symbol for equality and that includes the  $\circ$ -operator. The logic  $Qmbc_{\approx}$  is defined by a Hilbert Calculus that extends that for *mbc* over the signature  $\Theta$ , adding the following axiom schemas and inference rules:*

**Axiom schemas:**

$$\varphi[x/t] \rightarrow \exists x\varphi, \text{ if } t \text{ is a term free for } x \text{ in } \varphi \quad (\mathbf{Ax11})$$

$$\forall x\varphi \rightarrow \varphi[x/t], \text{ if } t \text{ is a term free for } x \text{ in } \varphi \quad (\mathbf{Ax12})$$

$$\alpha \rightarrow \beta, \text{ whenever } \alpha \text{ is a variant of } \beta \quad (\mathbf{Ax13})$$

$$\forall x(x \approx x) \quad (\mathbf{Ax14})$$

$$\forall x\forall y((x \approx y) \rightarrow (\alpha \rightarrow \alpha[x/y])), \text{ if } y \text{ is a variable free for } x \text{ in } \alpha \quad (\mathbf{Ax15})$$

**Inference rules:**

$$\frac{\alpha \rightarrow \beta}{\alpha \rightarrow \forall x\beta} \text{ if } x \text{ is not free in } \alpha \quad (\forall - \mathbf{In})$$

$$\frac{\alpha \rightarrow \beta}{\exists x\alpha \rightarrow \beta} \text{ if } x \text{ is not free in } \beta \quad (\exists - \mathbf{In})$$

**4.2 LFI-set theories.** The basic system of paraconsistent set theory that we discuss are based on the logic  $Qmbc_{\approx}$  and a first order signature  $\Theta_{ZF}$ , which contains, besides the equality predicate  $\approx$ , a binary predicate  $\in$  and a further unary predicate  $C$ , which stands for the consistency of sets—as opposed to formulas.

The possibility to distinguish between inconsistent and consistent sets is intended to capture a seminal intuition of Cantor, who made a clear separation between consistent and inconsistent totalities, that is, between sets and collections that are too big to be sets, like the totality of all alephs. This distinction was not only a terminological curiosity, since Cantor used the notion of inconsistent totality in a mathematical argument meant to show that the collection of cardinal is *totally* ordered by size; [13], p. 410.

DEFINITION 10. ([5]) *The system ZFmbC is the first order theory with equality obtained from the logic  $Qmbc_{\approx}$  over  $\Theta_{ZF}$  by adding the following set-theoretic axiom schemata:*

$$\begin{aligned} & \forall x\forall y[\forall z(z \in y \leftrightarrow z \in x) \rightarrow (x = y)] \text{ (Extensionality)} \\ & \forall x\exists y\forall z(z \in y \leftrightarrow \forall w \in z(w \in x)) \text{ (Power Set)} \\ & \forall x\exists y\forall z(z \in y \leftrightarrow \exists w \in x(z \in w)) \text{ (Union)} \\ & \forall x\forall y\exists z\forall w(w \in x \leftrightarrow (w \approx x \vee w \approx y)) \text{ (Pairing)} \\ & \exists w((\emptyset^* \in w) \wedge (\forall x)(x \in w \rightarrow x \cup \{x\} \in w)) \text{ (Infinity)}^8 \\ & FUN_{\varphi} \rightarrow \exists b\forall y(y \in b \leftrightarrow \exists x(x \in a \wedge \varphi(x, y))) \text{ (Replacement}_{\varphi})^9 \\ & (\exists b\forall x((x \in b) \leftrightarrow ((x \in a) \wedge \varphi(x)))) \text{ (Separation}_{\varphi}) \\ & C(x) \rightarrow (\exists y(y \in x) \rightarrow \exists((y \in x) \wedge \sim \exists z(z \in x \wedge z \in y))) \text{ (Weak regularity)} \\ & (x \not\approx y) \leftrightarrow \exists z((z \in x) \wedge (z \in y)) \vee \exists z((z \in y) \wedge (z \in x)) \text{ (Unextensionality)} \\ & \forall x(x \in y \rightarrow (C(x) \rightarrow C(y))) \text{ (Con}_0\text{)} \\ & \forall x(C(x) \rightarrow \circ(x = x)) \text{ (Con}_1\text{)} \\ & \forall x\neg \circ(x = x) \rightarrow \neg C(x) \text{ (Con}_2\text{)} \end{aligned}$$

Notice that in ZFmbC we have at disposal a classical  $\sim$  negation and a paraconsistent one  $\neg$ , since this is already the case in  $QmbC_{\approx}$ . Moreover, we can use  $x \not\approx y$  to abbreviate  $\neg(x \approx y)$  and  $x \notin y$  to abbreviate the formula  $\neg(x \in y)$ . Notice that the first six axioms,

<sup>8</sup> In the axiom of infinity  $\emptyset^*$  stands for the strong empty set  $\emptyset^* := \{x \mid \sim(x \approx x)\}$ .

<sup>9</sup> For the replacement schema let  $\varphi(x, y)$  be a formula where variables  $x$  and  $y$  occur free. Let  $z$  be a fresh variable. Then  $FUN_{\varphi}$  denotes the following formula:  $FUN_{\varphi} =_{Def} \forall x\forall y\forall z(\varphi(x, y) \wedge \varphi(x, z) \rightarrow (y \approx z))$ .

when added to classical first order logic with identity together with the regular regularity axiom constitute the ZF set theory. Therefore the non classical flavor of this set theory stems from Weak regularity, Unextensionality and the axioms governing the consistency predicate.

In [5] it is proved that ZFmbC is non-trivial provided that ZF is consistent. As a consequence, this paraconsistent set theory—as it is the case in  $V^{\text{PS}_3}$ —does not allow the Russel set in their respective ontology. On the other hand this set theory is extremely weak with regard to its expressive power, since we can not define inconsistent sets nor the inconsistency operator. To this aim, we need to consider extensions of ZmbC defined by taking stronger LFIs and appropriate axioms for the consistency predicate. The first system we consider is based on the logic mCi.

DEFINITION 11. ([5]) *The system ZFmCi is obtained from ZFmbC by adding the following axioms, for  $n \geq 0$  :*

$$\begin{aligned} \neg \circ \varphi &\rightarrow (\varphi \wedge \neg \varphi) \text{ (ci)} \\ \neg^{n+2} \circ \varphi &\rightarrow \neg^n \circ \varphi (\neg^n) \\ \forall x(\neg C(x) &\rightarrow \neg \circ (x \approx x)) \text{ (Con}_3\text{)} \\ \forall x(\neg C(x) &\rightarrow \neg \circ (x \in x)) \text{ (Con}_4\text{)} \end{aligned}$$

The first two axioms (ci) and  $(\neg^n)$  transform the underlying logic  $mbC$  into the stronger logic  $mCi$ , in which the inconsistency operator can be defined as  $\bullet\varphi := \neg \circ \varphi$  and inconsistent sets as the dual of consistent sets, so  $I(x) := \neg C(x)$ .

DEFINITION 12. *The system ZFCi is obtained from ZFmCi by adding axiom cf:*

$$\neg \neg \varphi \rightarrow \varphi \text{ (cf)}$$

DEFINITION 13. *Let ZFCil be the system obtained from ZFmCi by adding the following axioms:*

$$\begin{aligned} \neg(\varphi \wedge \neg \varphi) &\rightarrow \circ \varphi \text{ (cl)} \\ \forall x((\neg(x \approx x) \wedge (x \not\approx x)) &\rightarrow C(x)) \text{ (Con}_5\text{)} \\ \forall x((\neg(x \in x) \wedge (x \notin x)) &\rightarrow C(x)) \text{ (Con}_6\text{)} \end{aligned}$$

## 5. LFI-set theories and linear models

In this section we will show that the paraconsistent Zermelian set theories ZFmbC, ZFmCi, ZFCi and ZFCil are not valid in any  $\mathfrak{B}^*$ -valued model (we will refer to them simply as linear models). This result is counterintuitive, since by Theorem 4 the three-elements linear algebra is nothing else than  $(\text{PS}_3, *)$ , which is an LFI, as noted in [7]. This is odd and will reveal serious issues in the conception and motivation of these Zermelian set theories.

**5.1 ZF-axioms.** In order to check whether the axioms of an LFI-set theory is valid in a linear model, we first need to specify the interpretation of the  $\circ$ -operator. Since this operator is the source of classicality, we use the following evaluation  $\llbracket \circ \varphi \rrbracket = \llbracket \varphi \rightarrow \perp \rrbracket \vee \llbracket \sim \varphi \rightarrow \perp \rrbracket$ , which expresses the fact that  $\varphi$  is consistent if either  $\varphi$  or its negation is false. In the three element case this gives rise to the following table.

$\varphi$	$\circ\varphi$
1	1
$\frac{1}{2}$	0
0	1

Figure 3. Three valued matrix for consistency operator

We now check the validity of the axioms of ZFmbC in the linear models. Since the focus of this work is paraconsistency we require our models to be non-classical.

DEFINITION 14. *By a paraconsistent linear algebra we mean a  $\mathfrak{V}$ -algebra  $\mathbb{L}$  such that  $|\mathbb{L}| > 2$ . The class of such algebras will be indicated by  $\mathfrak{V}^*$  and its elements  $\mathfrak{V}^*$ -algebras. By a paraconsistent linear model we mean a  $\mathfrak{V}$ -model of the form  $\mathbf{V}^{\mathbb{L}}$ , with  $|\mathbb{L}| > 2$ , and we refer to it as a  $\mathfrak{V}^*$ -model.*

By Theorem 4 we know that the negation free fragment of ZF is valid in any linear model. Therefore, any linear model verifies (*Union*), (*Extensionality*), (*Power set*), (*Separation $_{\varphi}$* ), (*Foundation $_{\varphi}$* )<sup>10</sup>, and (*Replacement $_{\varphi}$* ), for  $\varphi$  negation free.

On the other hand, (*Infinity*) and (*Weak regularity*), as presented in [4], are not valid in any  $\mathfrak{V}^*$ -model. Nonetheless, by a suitable modification we can obtain their validity, without distorting the spirit of ZFmbC. The issue with (*Infinity*) consists in the use of the *strong* empty set  $\emptyset^*$  which cannot be defined in the  $\mathfrak{V}^*$ -models. In [4], the authors define the strong empty set using (*Separation*) and (*Extensionality*) and the fact that  $\text{ZFmbC} \vdash ((x \in a) \wedge \sim (x = x)) \leftrightarrow \sim (x = x)$ . However, this cannot be done in any  $\mathfrak{V}^*$ -model, since the formula that defines the strong empty set is not a negation free instantiation of separation. But this is not dramatic, since we can use a different, negation free, formulation of (*Infinity*) that suits better an algebra-valued model treatment of set theory:<sup>11</sup>

$$\exists x[\exists y(\forall z(z \in y \rightarrow \perp) \wedge y \in x) \wedge \forall w \in x \exists u \in x(w \in u)]$$

On the other hand, (*Weak regularity*) states that only consistent sets cannot be sets of themselves. But notice that (*Foundation $_{\varphi}$* ) is valid in any  $\mathfrak{V}^*$ -model, for  $\varphi$  negation free and therefore we have that  $\llbracket x \in x \rrbracket^{\mathbb{L}} = 0$ , for all  $x \in \mathbf{V}^{\mathbb{L}}$  and  $\mathbb{L} \in \mathfrak{V}$ . Thus even inconsistent sets cannot be ill-founded. This seems to suggest that we should discard (*Weak regularity*) in favor of (*Foundation $_{\varphi}$* ), for  $\varphi$  negation free.

For what concerns (*Unextensionality*) the situation is more serious. Before stating the relevant result, a word on notation. When we say that an algebra-valued model satisfies a formula we mean that, for any filter we can define on the algebra, the evaluation will yields a value in the set of designated values. In this case we write  $\mathbf{V}^{\mathbb{L}} \models \varphi$ . On the contrary, if we want to stress the selected filter we use the following notation. Given a filter  $G \subseteq A$  defined on the base set of an algebra  $\mathbb{A}$  we write  $\mathbf{V}^{\mathbb{L}} \models_G \varphi$  to say that  $\varphi$  is valid in  $\mathbf{V}^{\mathbb{L}}$  according to the consequence relation engendered by the filter  $G$ .

THEOREM 15. *Let  $\mathbf{V}^{\mathbb{L}}$  be a  $\mathfrak{V}^*$ -model, then  $\mathbf{V}^{\mathbb{L}} \not\models (\text{Unextensionality})$ .*

PROOF. We prove that  $\mathbf{V}^{\mathbb{L}} \not\models (\text{Unextensionality})$  by showing that there is a filter  $G \subseteq L$  such that  $\mathbf{V}^{\mathbb{L}} \not\models_G (\text{Unextensionality})$ . Notice that  $(\text{PS}_3, *)$  is contained in any

<sup>10</sup> Foundation is here intended in terms of transfinite recursion, which is indeed a schema

<sup>11</sup> This definition of the axiom of infinity was already proposed in [1].

$\mathfrak{V}^*$ -algebra. Let us call the elements of the base set of  $(\mathbb{P}\mathbb{S}_3, *)$  as  $\{0, \frac{1}{2}, 1\}$ . And let define  $G$  to be a filter that contains  $\frac{1}{2}$ , and hence all the elements above  $\frac{1}{2}$ .

We can consider the following names as witnesses for all  $\mathbf{V}^*$ -models:  $p_{\frac{1}{2}} = \langle \emptyset, \frac{1}{2} \rangle$ ,  $x = \langle p_{\frac{1}{2}}, 1 \rangle$ ,  $y = \langle p_{\frac{1}{2}}, \frac{1}{2} \rangle$ . Now, we simply calculate the value of (*Unextensionality*) in a generic  $\mathfrak{V}^*$ -model  $\mathbf{V}^{\mathbb{L}}$ :

$$\begin{aligned} &= \bigvee_{w \in \mathbf{V}^{\mathbb{L}}} (\llbracket w \in x \rrbracket \wedge \llbracket w \in y \rrbracket) \vee \bigvee_{w \in \mathbf{V}^{\mathbb{L}}} (\llbracket w \in x \rrbracket \wedge \llbracket w \in y \rrbracket) \\ &= (1 \wedge \frac{1}{2}) \vee (1 \wedge \frac{1}{2}) \end{aligned}$$

Now we can deduce the truth values of  $\llbracket w \in y \rrbracket^* = \frac{1}{2}$  and  $\llbracket w \in x \rrbracket^* = 0$ . Furthermore, it is easy to show that  $\llbracket x = y \rrbracket = 1$  and by Lemma 6  $\llbracket \neg(x = y) \rrbracket = \llbracket x = y \rrbracket^* = 0$ . Now, we can calculate readily

$$\bigvee_{w \in \mathbf{V}^{\mathbb{L}}} (\llbracket w \in x \rrbracket \wedge \llbracket w \in y \rrbracket^*) \vee \bigvee_{w \in \mathbf{V}^{\mathbb{L}}} (\llbracket w \in x \rrbracket^* \wedge \llbracket w \in y \rrbracket) = \frac{1}{2}$$

Since we know that the antecedent of the unextensionality axiom receives truth value  $\frac{1}{2}$  and the consequent truth value 0, we can conclude by Corollary 3.11 that the right-to-left implication of (*Unextensionality*) and therefore the axiom itself, receives truth value 0.

Notice that the above argument can be adapted to show that, given a  $\mathfrak{V}^*$ -model  $\mathbf{V}^{\mathbb{L}}$  and a filter  $G \subseteq L$  such that  $|G| > 2$ , there are always three elements in  $\mathbf{V}^{\mathbb{L}}$  that witness the failure of (*Unextensionality*). It is sufficient to choose the relevant intermediate element of  $\mathbb{L}$ —that plays the role of  $\frac{1}{2}$  in  $(\mathbb{P}\mathbb{S}_3, *)$ —high enough to be in the filter.<sup>12</sup> The case  $|G| = 1$  is not relevant since in this case  $G = \{1\}$  and therefore the resulting consequence relation is classical.

Therefore (*Unextensionality*) fails very badly in any  $\mathfrak{V}^*$ -model, disregarding the choice of the consequence relation. This is already troubling, but a possible strategy would consist in biting the bullet and just to eliminate this axioms. However, we will show that the issues of LFI-set theories lay much deeper in the interpretation of inconsistent sets.

**5.2 C-axioms.** In [5] the definition of inconsistency is given in terms of the violation of the basic logical operations of set theory: equality and membership.

$$(i) \neg C(x) \leftrightarrow (x \in x)$$

$$(ii) \neg C(x) \leftrightarrow \neg \circ (x \in x)$$

$$(iii) \neg C(x) \leftrightarrow (x \not\approx x)$$

$$(iv) \neg C(x) \leftrightarrow \neg \circ (x \not\approx x)$$

We now show that these definitions cannot capture the notion of inconsistency in any linear model.

<sup>12</sup> Notice that it is in this part of the argument that we use Lemma 6, to show that an intermediate value is mapped to itself by the  $*$ -negation.

PROPOSITION 16. For  $\mathbb{L} \in \mathfrak{V}^*$ , we have that (i)  $\llbracket x \in x \rrbracket^{\mathbb{L}} = 0$ , (ii)  $\llbracket \neg \circ (x \in x) \rrbracket^{\mathbb{L}} = 0$ , (iii)  $\llbracket x \neq x \rrbracket^{\mathbb{L}} = 0$  and (iv)  $\llbracket \neg \circ (x \neq x) \rrbracket^{\mathbb{L}} = 0$ .

PROOF. Notice that we have already argued that (*Foundation* <sub>$\varphi$</sub> ), for  $\varphi$  negation free, yields  $\llbracket x \in x \rrbracket^{\mathbb{L}} = 0$ . This establishes (i). From (i) we know that  $\llbracket x \in x \rrbracket^{\mathbb{L}}$  receives a classical truth value. Hence we deduce  $\llbracket \circ (x \in x) \rrbracket^{\mathbb{L}} = 1$ . Then by Lemma 6  $\llbracket \neg \circ (x \in x) \rrbracket^{\mathbb{L}} = 0$ , establishing (ii). Adapting Proposition 1 of [9], we know that for every  $x \in \mathbf{V}^{\mathbb{L}}$   $\llbracket x \approx x \rrbracket^{\mathbb{L}} = 1$  and  $u(x) \leq \llbracket x \in u \rrbracket$  for every  $u$  in the domain of  $x$ , so  $\llbracket x \neq x \rrbracket^{\mathbb{L}} = 0$ . From this follows (iii). Furthermore, given the classical truth value of  $\llbracket (x \neq x) \rrbracket^{\mathbb{L}}$  it follows that  $\llbracket \circ (x \neq x) \rrbracket^{\mathbb{L}} = 1$ . Finally, given Lemma 6 we can conclude  $\llbracket \neg \circ (x \neq x) \rrbracket^{\mathbb{L}} = 0$ , which settles (iv).

This proposition shows clearly that none of the above interpretations of inconsistency can ever be realized in any  $\mathfrak{V}^*$ -model.

## 6. Conclusion

The failure of the characteristic axioms of these LFI-set theories, together with the impossibility to define inconsistency in the way proposed in [5], leaves us with two options. Either we assume that,  $\forall x \in \mathbf{V}^{\mathbb{L}} \llbracket \neg C(x) \rrbracket^{\mathbb{L}} = 0$ , therefore assuming that every set is consistent, or we can try to modify the interpretation of the consistency predicate in order to provide examples of inconsistent sets in  $\mathfrak{V}^*$ -models. We argue that both strategies are unfeasible.

Let us make explicit the two horns of this dilemma:

**Horn 1:** The first strategy consists in eliminating inconsistent totalities from the picture. This option is, as a matter of fact, not very far from Cantor's ideas. Indeed in several letters to Hilbert and Jourdain ([13], p.425–435) Cantor identified sets and consistent totalities, arguing that the universe of set was the collection of all totalities that did not lead to any contradiction. On this basis we can therefore postulate an axiom, that we can call *Cantorian Axiom*, that expresses this intuition.

$$C(x) \leftrightarrow \text{Set}(x) \quad (\text{Cantorian Axiom})$$

Now, if we extend ZFmbC with the Cantorian Axiom we obtain that all  $\mathfrak{V}$ -models validate this set theory and all its extension. Indeed these LFI-set theories would all be equal ZF, since we have just removed all inconsistent sets. However, this move would totally trivialize the main motivation of using a consistency predicate in dealing with an inconsistent set theory, as clearly expressed [4], p. 366.

The main idea is to assume that not only sentences can be taken to be consistent or inconsistent, but also that sets themselves can be thought to be consistent or inconsistent. We establish the basis for new paraconsistent set-theories (such as ZFmbC and ZFCil) under this perspective and establish their non-triviality, provided that ZF is consistent.

**Horn 2:** On the other hand, we could try to modify the interpretation of the (in)consistency predicate, in order to have at least one set  $x$  such that  $\llbracket \neg C(x) \rrbracket \geq a$ , for a designated value  $a \neq 1$  of a  $\mathfrak{V}^*$ -algebra. However, also this strategy is doomed to fail, since no  $\mathfrak{V}^*$ -model would be able to validate ZFmCi, or ZFCi, or ZFCil.

LEMMA 17. Let  $\mathbf{V}^{\mathbb{L}}$  be a  $\mathfrak{V}^*$ -model,  $G \subseteq L$  a filter containing an element  $a \in L$  that is different from 1, and  $x \in \mathbf{V}^{\mathbb{L}}$  such that  $\llbracket \neg C(x) \rrbracket^{\mathbb{L}} \geq a$ . Then  $\mathbf{V}^{\mathbb{L}} \not\models \neg C(x) \rightarrow \neg \circ (x = x)$



PROOF. Due to Proposition 1 in [9], we know that  $\llbracket \circ(x = x) \rrbracket^{\mathbb{L}} = 1$  and therefore  $\llbracket \neg \circ(x = x) \rrbracket^{\mathbb{L}} = 0$ . Then by Corollary 3.11 :

$$\llbracket \neg C(u) \rrbracket^{\mathbb{L}} \geq a \rightsquigarrow \llbracket \neg \circ(u = u) \rrbracket^{\mathbb{L}} = 0.$$

Notice that in this case we would validate (vacuously) all the axioms of ZFmbC that govern the consistency predicate, with exception of **con**<sub>0</sub>. This is due to the fact that either  $\neg C(x)$  occurs in the consequent of an axiom, so that the antecedent is always false, as it is the case with **con**<sub>2</sub>, or  $C(x)$  occurs in the antecedent, so that the consequent will be trivially true, as in **con**<sub>1</sub>.

Besides the formal problems that arise from accepting the existence of inconsistent totalities, there is also a deeper conceptual issue. A charitable interpretation of the inconsistency predicate—that seems faithful to the history of set theory—would consist in equating inconsistent objects and proper classes. However, at a closer look the axioms of ZFmbC force us to discard this possibility, since the axiom **con**<sub>0</sub> would imply that the universal class is a set and not a proper class. Indeed, it says that the property of being a set is propagated  $\in$ -upward, which clearly cannot be the case since all elements of  $\mathbf{V}$  are, by definitions, sets.

We can therefore conclude, from **Horn 1** and **Horn 2**, that the LFI-set theory we discussed are not valid in the most natural models for Zermelian set theories at disposal. This is even more belittling, considering that the internal logic of  $\mathbf{V}^{(\mathbb{P}\mathbb{S}_3, *)}$  is an LFI. Not only we need to give up the characteristic axiom (*Unextensionality*), but **Horn 2**, together with the charitable interpretation of inconsistent totalities as classes, shows that not even the weak ZFmbC-axioms that govern the  $C$ -predicate can hold in a  $\mathfrak{V}^*$ -model. We believe that the fact that inconsistent totalities do not find a place in these LFI-set theories is a serious betrayal of their original motivation, which therefore suggest a fresh new start. The failure of this attempt leaves open the challenging problem of making inconsistent totalities compatible with a Zermelian set theory based on a paraconsistent logic. We leave this open question for future work.

### Acknowledgements

The first author acknowledges support from the FAPESP grant. n. 2017/23853-0. The second author acknowledges support from the FAPESP Jovem Pesquisador grant n. 2016/25891-3.

## Bibliography

- [1] J. Bell. *Set theory. Boolean valued models and independence proofs*. Oxford Science Publications, 2005.
- [2] George Boolos. The Iterative Conception of Set. *The Journal of Philosophy*, 68(8):215–231, April 1971.
- [3] Ross T. Brady. The non-triviality of dialectical set theory. In G. Priest, R. Routley, and J. Norman, editors, *Paraconsistent Logic: Essays on the Inconsistent*, pages 437–470. Philosophia Verlag, 1989.
- [4] Walter Carnielli and Marcelo E. Coniglio. *Paraconsistent Logic: Consistency, Contradiction and Negation*. Basel, Switzerland: Springer International Publishing, 2016.
- [5] Walter Carnielli and Marcelo E. Coniglio. Paraconsistent set theory by predicating on consistency. *Logique Et Analyse*, 26(1):97–116, 2016.
- [6] Paul Cohen. The independence of the continuum hypothesis. *Proc. Nat. Acad. Sci. USA*, 50(6):1143–1148, 1963.
- [7] M. E. Coniglio and L. H. D. C. Silvestrini. An alternative approach for quasi-truth. *Logic Journal of the IGPL*, 22(2):387–410, 2014.
- [8] Newton Ca da Costa. On paraconsistent set theory. *Logique Et Analyse*, 115:361–371, 1986.
- [9] Benedikt Löwe and Sourav Tarafder. Generalized algebra-valued models of set theory. *Review of Symbolic Logic*, 8(1):192–205, 2015.
- [10] João Marcos. 8K solutions and semi-solutions to a problem of da Costa. *Unpublished*.
- [11] João Marcos. Logics of Formal Inconsistency. *PhD Thesis*.
- [12] Santiago J. Martinez and Giorgio Venturi. Non-classical Models of ZF. *Studia Logica*, 2020. DOI: <https://doi.org/10.1007/s11225-020-09915-0>. (Updated on July 11, 2020).
- [13] W. Nilson and H. Meschkowski. *Georg Cantor. Briefe*. Springer Verlag, 1991.
- [14] Greg Restall. A note on naive set theory in  $\{\text{Rm LP}\}$ . *Notre Dame Journal of Formal Logic*, 33(3):422–432, 1992.
- [15] Zach Weber. Notes on inconsistent set theory. In Francesco Berto, Edwin Mares, Koji Tanaka, and Francesco Paoli, editors, *Paraconsistency: Logic and Applications*, pages 315–328. Springer, 2013.

# HESITAÇÕES NA ARITMÉTICA DE FREGE

Fernando Ferreira <sup>1</sup>

Em 1884, Gottlob Frege fecha a introdução d’*Os Fundamentos da Aritmética* com a formulação de três princípios fundamentais:

- (1) É necessário separar com nitidez o que é psicológico do que é lógico, o que é objetivo do que é subjetivo.
- (2) Só se pode perguntar pela denotação de uma palavra no contexto de uma proposição e não considerando-a isoladamente.
- (3) Deve manter-se sempre presente a distinção entre conceito e objeto.

Estas três injunções metodológicas estão na base do programa logicista de Frege de redução da aritmética à lógica. Estão mesmo na base do surgimento da filosofia analítica.

O volume d’*Os Fundamentos da Aritmética* é hoje considerado uma obra-prima da filosofia. Contém, numa primeira parte, críticas às opiniões de filósofos e matemáticos da época acerca da noção de número e termina com a exposição duma nova tese para a fundamentação da aritmética. Não obstante, esta é uma obra essencialmente programática e pouco amadurecida. O próprio Frege afirma n’*Os Fundamentos da Aritmética* que não fez nada “mais do que ter tornado verosímil a natureza analítica das proposições aritméticas” e adianta que dúvidas quanto à natureza puramente lógica dos raciocínios expostos apenas podem ser dissipadas “por meio de uma apresentação de uma cadeia dedutiva sem lacunas”. O *magnum opus* de Frege são os dois tomos das *Grundgesetze der Arithmetik* (Leis Fundamentais da Aritmética), publicados com um intervalo de dez anos, em 1893 e 1903. Nesta obra, Frege leva a cabo o programa logicista nos termos com os quais se achava comprometido, i.e., por meio de deduções estritamente formais. Para o fazer, Frege socorre-se do seu anterior trabalho *Begriffsschrift* (Notação Conceptual) de 1879. Na nossa opinião, a grandeza imorredoura de Frege encontra-se precisamente nesta obra precoce. De uma assentada, Frege faz nesta obra duas contribuições geniais. Em primeiro lugar, inventa uma linguagem formal para a quantificação, em que os quantificadores podem aparecer encadeados. Depois, oferece-nos um cálculo formal para efetuar deduções nesta linguagem (o cálculo de predicados).

É bem certo que se pode pôr em causa a importância que Frege dá a formalizar completamente os raciocínios da sua dedução lógica da aritmética. Hoje em dia, muitos argumentarão que isso não é essencial e que, de facto, é a própria informalidade d’*Os Fundamentos da Aritmética* que a torna uma obra acessível, conhecida e lida. Também eu partilho desta opinião. Porém, não se deve passar em claro as enormes consequências involuntárias que a opção fregiana teve. As linguagens de programação são linguagens formais e podem ser consideradas herdeiras afastadas da visão fregiana. Mais próximo das preocupações de Frege, o *Begriffsschrift* é o primeiro sistema formal dedutivo depois do cálculo silogístico aristotélico, ultrapassando-o vastamente em poder expressivo. Com

---

<sup>1</sup>Departamento de Matemática, Faculdade de Ciências, Universidade de Lisboa. Email: [fjferreira@fc.ul.pt](mailto:fjferreira@fc.ul.pt).

---

efeito, o cálculo de predicados é capaz de formalizar axiomáticamente a prática matemática, o que deu origem à disciplina da Lógica Matemática e aos estudos metamatemáticos. No âmbito da Lógica Matemática estas formalizações ocorrem, em grande parte, como formalizações *em princípio*. Porém, com o aparecimento recente dos *assistentes de demonstrações* (“proof assistants”), a visão fregiana do rigor, no seu sentido mais estrito, encontrou o seu herdeiro natural e do qual – tenho essa convicção – ainda se vai ouvir falar muito nas próximas décadas.

Afirmei acima que *Os Fundamentos da Aritmética* é uma obra pouco amadurecida. Tal não obsta a que se encontre n’*Os Fundamentos da Aritmética* grandes ideias e análises, de que são exemplos a defesa de que o número se aplica a um conceito ou o esboço da dedução lógica das leis fundamentais da aritmética a partir do princípio de Hume. Porém, em certas alturas cruciais, a obra é surpreendentemente hesitante e ambígua, oscilando fortemente entre posições incompatíveis. Frege acaba por não fazer uso direto da segunda injunção para explicar a apreensão de números, ainda que esta injunção esteja sempre presente como um espectro no logicismo d’*Os Fundamentos da Aritmética*. Por outro lado, se bem que a terceira injunção seja sempre mantida na forma, ela não é mantida na substância e isso veio a revelar-se fatal por causa do paradoxo de Russell.

Gostaria de terminar este introito com uma observação pouco habitual. Um leitor de Frege, mesmo que não muito atento, não deixará de notar o modo mordaz com que por vezes Frege critica posições rivais. Esta forma de proceder deixou marca nefasta nalguma prática da filosofia analítica, traduzindo-se ocasionalmente por certas formas pouco corteses de criticar posições adversárias. Também, aqui e ali, levou a um estilo de filosofar muito apoiado no *bon mot* em prejuízo duma sobriedade distanciada. Não está em causa o facto de Frege ter razão na grande maioria das críticas que fazia, mas o tom e a insistência com que as fazia provavelmente gerava algum desconforto e, certamente, não ajudou a que o seu trabalho fosse justa e devidamente apreciado mais cedo.

## II

N’*Os Fundamentos da Aritmética*, Frege propõe uma nova forma de explicar o acesso epistemológico a objetos abstratos. Tomemos o próprio exemplo de Frege da direção duma reta. Dada uma reta  $\ell$ , o termo “a direção da reta  $\ell$ ” denota um objeto autónomo. Isto não requer, no entanto, que se tenha de dar uma denotação a este termo quando ele é considerado isoladamente. De acordo com a segunda injunção fregiana, apenas temos de responder pela denotação duma expressão no contexto duma proposição. Frege elege as igualdades entre direções como caso paradigmático destas proposições. De acordo com Frege, ao utilizarmos termos como “a direção da reta  $\ell$ ”, temos de estar na posse dum critério que permita afiançar se esse termo denota o mesmo objeto que outro termo do mesmo género. É claro que possuímos esse critério e ele é o seguinte: a direção da reta  $\ell$  é igual à direção da reta  $r$  se, e somente se,  $\ell$  e  $r$  são retas paralelas.

O critério acima estabelece, nas palavras de Frege, “o conteúdo de um juízo de reconhecimento” e marca uma forma de abstracionismo que se harmoniza muito naturalmente com as três injunções fregianas. A segunda injunção, como vimos, desempenha um papel crucial na própria formulação deste abstracionismo. Por outro lado, ao evitar considerar expressões do género “a direção da reta  $\ell$ ”, tomadas isoladamente, não caímos na tentação de invocar representações subjetivas de direções ou de apelar a uma intuição de direções. O critério fregiano de igualdade é puramente objetivo. A primeira injunção é, portanto,

---

cumprida escrupulosamente. Finalmente, Frege não define a direção da reta  $\ell$  como sendo o *conceito* “ser uma reta paralela à reta  $\ell$ ”. Para Frege, as direções são objetos e (de acordo com a terceira injunção) devemos ter sempre presente a distinção entre conceito e objeto. Uma direção não é um conceito.

Todavia, a situação da teoria abstracionista de Frege n’ *Os Fundamentos da Aritmética* é ambígua e problemática. Ainda que Frege chame com frequência a atenção para os juízos de reconhecimento, acaba por não subscrever o abstracionismo que os torna tão centrais. Diria que Frege *gosta* deste abstracionismo, que o acha uma muito boa ideia, mas que – finalmente – renuncia a ele a contragosto. Renuncia a ele umas poucas páginas depois de o ter introduzido. A (auto)crítica é contundente. Não basta ter maneira de determinar quando é que as igualdades da forma “a direção da reta  $\ell$  é igual à direção da reta  $r$ ” são verdadeiras para que, automaticamente, tenhamos um método para determinar o valor de verdade de todas as proposições em que ocorram termos do género “a direção da reta  $\ell$ ”. É também necessário que se dê um método para determinar o valor de verdade das proposições da forma “a direção da reta  $\ell$  é igual a  $t$ ”, onde  $t$  é um termo (denotando um objeto) não necessariamente do género “a direção da reta  $r$ ”. Frege põe a questão da seguinte maneira: como é que sabemos “se a Inglaterra é o mesmo que a direção do eixo da Terra”? A pergunta parece estapafúrdia mas, de facto, revela uma grande finura lógico/filosófica por parte de Frege. Ninguém iria confundir a Inglaterra com a direção do eixo a Terra, mas a distinção não se explica a partir do critério da igualdade de direções. “Falta-nos o conceito de direção”, afirma Frege e, é claro, anda-se em círculo se dissermos que  $t$  é uma direção se  $t$  for a direção dalguma reta. Este problema é hoje conhecido por problema do Júlio César pois, n’ *Os Fundamentos da Aritmética*, Frege pergunta “se este conhecido conquistador da Gália é ou não um número”.

Chegado a este ponto, Frege tenta outro caminho. Define agora as direções explicitamente (e não através de proposições de reconhecimento): a direção da reta  $\ell$  é a extensão do conceito “ser uma reta paralela à reta  $\ell$ ”. Frege renuncia aqui ao uso direto da segunda injunção. Agora já se sabe qual é o sentido da expressão “a direção da reta  $\ell$ ”, tomada isoladamente: é a extensão do conceito “ser uma reta paralela à reta  $\ell$ ”. Frege não se alonga sobre a noção de extensão, apenas a discutindo por duas ou três vezes n’ *Os Fundamentos da Aritmética*. Diz primeiramente, numa nota de pé de página, que extensões são objetos e que pressupõe que saibamos o que é uma extensão. Esta suposição não é discutida n’ *Os Fundamentos de Aritmética* podendo-se, contudo, imaginar que Frege apenas dá sentido a expressões do género “a extensão do conceito  $F$ ” através da segunda injunção. Talvez seja esse o caso, como discutiremos mais tarde.

Quando se lê *Os Fundamentos da Aritmética*, fica-se com a sensação de que a solução encontrada – por meio de extensões – não satisfaz Frege plenamente. É instrutivo citar Frege mais longamente, agora na última vez em que discute a noção de extensão nessa obra: “[...] pressupusemos como conhecido o sentido da expressão ‘extensão do conceito’. Esta maneira de ultrapassar a dificuldade [...] não encontrará, é certo, aplauso unânime e muitos preferirão [...] outro processo. Eu também não atribuo qualquer peso decisivo à participação nela da extensão de um conceito”. A posição de Frege é vacilante. Talvez isso explique por que razão é que, na nota de pé de página atrás referida, Frege sugere – com uma enorme ligeireza – um caminho alternativo, sem extensões, em que propõe que se considere que a direção da reta  $\ell$  seja o *conceito* “ser uma reta paralela à reta  $\ell$ ” (de facto, esta proposta é feita para números – não direções – mas isto é incidental). Frege

---

faz uma afirmação destas em flagrante contradição com o que defendia pouco antes. Se bem que seja verdade que Frege lembre nessa mesma nota de pé de página as objeções contra tomar direções (e números) como conceitos, acaba por afirmar candidamente que “embora eu seja da opinião de que [...] as objeções podem ser refutadas, isso iria aqui levar-nos demasiado longe”. Esta é, talvez, a nota de pé de página mais momentosa de toda a história da filosofia analítica ...

### III

Para Frege, o problema fundamental da aritmética é este: “como é que apreendemos objetos lógicos e, em particular, os números?” Acima, focámo-nos em direções em vez de números pois isso facilitou a discussão de certas questões. Numa recapitulação do que foi feito n’*Os Fundamentos da Aritmética*, Frege escreve que “só seguindo este princípio [segunda injunção], penso eu, se pode evitar a conceção fisicalista do número sem se cair na conceção psicológica do mesmo”. Para Frege, os números aplicam-se a conceitos. Por exemplo, o número 1 aplica-se ao conceito “ser satélite natural da Terra”. Noutra maneira de dizer: para o conceito “ser satélite natural da Terra” vem o número 1. A cada conceito está associado o número que vem para ele, tal como a cada reta está associada a sua direção. A expressão “o número que vem para o conceito  $F$ ” denota um objeto mas tal não requer que se tenha que dar uma denotação a este termo quando ele aparece isoladamente. A denotação destes termos é, de acordo com Frege, dada através do contexto de juízos de reconhecimento. O critério de igualdade para os números é conhecido como o princípio de Hume: o número que vem para o conceito  $F$  é igual ao número que vem para o conceito  $G$  se, e somente se, os conceitos  $F$  e  $G$  são equinumerícos. Neste critério, a noção de equinumericidade entre conceitos é a noção, hoje familiar, da existência duma bijeção entre os objetos que caem sob  $F$  e aqueles que caem sob  $G$ .

Como se disse na secção anterior, Frege renuncia a esta forma de abstracionismo por causa do problema do Júlio César. Porém, há algo que parece ter escapado a Frege pois há um problema adicional com este abstracionismo. Há uma circularidade subtil no critério de reconhecimento para números dado pelo princípio de Hume. Com efeito, na formulação da equinumericidade de conceitos ocorre o sinal de igualdade: ocorre, obviamente, na própria formulação do que é ser uma bijeção, mas também pode ocorrer na expressão sintática dos conceitos em causa. Ora, se estivermos a contar números, p. ex., a considerar a igualdade numérica entre o número que vem para o conceito “ser um número primo menor do que 15” e o número que vem para o conceito “ser um divisor de 12”, a formulação de que estes dois conceitos são equinumerícos (a de que há uma bijeção entre os objetos – números – que caem sob um e outro conceito) envolve igualdades entre números. Ora estas igualdades, *pace* o problema do Júlio César, devem ser explicadas pelo critério de igualdade que advém do princípio de Hume. Há aqui um círculo que não existe no caso das direções.

Uma forma de evitar esta circularidade seria impor que somente se façam contagens de objetos físicos, ou seja, que apenas se considerem equinumericidades entre conceitos que se aplicam a objetos físicos. Esta seria, claramente, uma solução *ad hoc* pois para Frege todos os objetos podem ser contados, inclusive os próprios números. Mas o caso é ainda mais sério pois esta solução não é adequada aos propósitos de Frege: se não for permitido contar números, o desenvolvimento fregiano da aritmética fica tecnicamente impossível de fazer.

---

N’*Os Fundamentos da Aritmética*, Frege acaba por dar a seguinte definição explícita de número: o número que vem para o conceito  $F$  é a extensão do conceito “ser equinúmero ao conceito  $F$ ”. É de observar que esta definição mostra, uma vez mais, que a posição de Frege n’*Os Fundamentos da Aritmética* ainda não estava totalmente amadurecida. Com efeito, Frege fala dum extensão cujos elementos são *conceitos*, nomeadamente aqueles conceitos que são equinúmericos ao conceito dado  $F$ . Mais tarde, nas *Grundgesetze der Arithmetik*, Frege já não se exprime assim. Nesta obra, as extensões são sempre extensões constituídas por objetos. O número que vem para o conceito  $F$  é agora definido como a extensão do conceito “ser uma extensão dum conceito equinúmero ao conceito  $F$ ”.

#### IV

Nas *Grundgesetze der Arithmetik* deixa de haver hesitações sobre qual é o motor do logicismo. É, notoriamente, a existência de cursos de valores (*Werthverlaufe*) de funções. Frege é cristalino no prefácio das *Grundgesetze der Arithmetik*: “sem [cursos de valores] não conseguiríamos ser bem sucedidos”.

Frege formula a sua famosa lei básica (V): o curso de valores da função  $f$  é igual ao curso de valores da função  $g$  se, e somente se,  $f$  e  $g$  têm o mesmo valor para todo os argumentos (objetos). As extensões de conceitos são um caso particular de cursos de valores pois, para Frege, um conceito é tão somente uma função cujos valores são o verdadeiro e o falso (que, para Frege, são dois objetos distintos). Evidentemente, a lei básica (V) tem a mesma forma do princípio de Hume. Ora, está vedado a Frege justificar a lei básica (V) como um princípio de abstração pois isso seria simplesmente deslocar o problema do Júlio César dos números para os cursos de valores (extensões).

Frege está ciente do problema de que a lei básica (V) não determina a denotação dos cursos de valores e levanta a questão “Agora, como é que se resolve esta indeterminação?” Há uma tentativa de responder a esta interrogação nas *Grundgesetze der Arithmetik* mas – de modo nada característico para Frege – é muito pouco claro aquilo que é feito. Suponho que se tenta fixar o valor de verdade das proposições exprimíveis no sistema formal das *Grundgesetze der Arithmetik*. Em termos fregianos, tenta-se fixar a denotação daqueles nomes próprios que referem o verdadeiro ou o falso (nomes que exprimem pensamentos, na terminologia de Frege). Dada a ênfase na determinação (dos valores de verdade) destes nomes próprios de carácter proposicional, reconhece-se (talvez) nesta estratégia a segunda injunção. Contudo, a tentativa é defeituosa. Ela falha. O problema não está (somente) numa sub-determinação das denotações das expressões. Jaz, desgraçadamente, também numa sobre-determinação: a denotação da expressão que corresponde à proposição paradoxal russelliana é, simultaneamente, o verdadeiro e o falso.

Há pelo menos duas passagens nas *Grundgesetze der Arithmetik* em que Frege manifestou uma preocupação com a existência de cursos de valores. Numa passagem tardia, já no segundo volume, a discussão ocorre no âmbito dum crítica de Frege às denominadas definições criativas. Para Frege, o matemático nada cria e, neste aspecto, é como o geógrafo, limitando-se a mapear o que existe. Frege já o tinha referido n’*Os Fundamentos da Aritmética* e é instrutivo citar esta passagem: “O matemático, tal como o geógrafo, não pode criar arbitrariamente; também ele não pode senão descobrir o que já existe e atribuir-lhe um nome”. Contra isto “pode ser apontado que [...] nós próprios criámos objetos, nomeadamente os cursos de valores. O que é que nós, de facto, fizemos aqui?” Imediatamente de seguida, Frege reformula a pergunta pela negativa: “Ou, para

---

começar: o que é que nós não fizemos? Não fizemos uma lista de propriedades para depois dizer: criamos uma coisa que verifica estas propriedades”. De acordo com Frege, o que fizemos foi reconhecer que há qualquer coisa em comum entre duas funções que tomam os mesmos valores para cada o objeto e chamar a esse algo em comum o curso de valores das duas funções. O direito de proceder assim é-nos dado, segundo Frege, por uma lei básica da lógica: o direito de converter a generalidade numa igualdade (identidade). Por outras palavras, o direito a converter a generalidade que exprime a igualdade dos valores tomados por duas funções, numa identidade entre objetos (entre os cursos de valores dessas funções). Frege insiste: “Podemos chamar de criação a este procedimento? A discussão desta questão pode facilmente degenerar numa alteração sobre palavras”. Seja como for, conclui Frege, a existência de cursos de valores incorpora, através dum único princípio (e numa vez por todas), o que é necessário para erigir o material das *Grundgesetze der Arithmetik*. Em vez de dar justificações *ad hoc* da existência de entes matemáticos sempre que deles necessitamos, com os cursos de valores estamos na posse dum procedimento universal. O leitor coevo há-de reconhecer aqui uma precoce declaração do papel fundamental da teoria dos conjuntos na matemática como tribunal da existência de entes matemáticos.

A primeira passagem em que Frege discute a existência de cursos de valores, é uma variação deste tema da universalidade. A existência de entes matemáticos converge para um só ponto: a existência de cursos de valores. Por isso “tanto quanto me é dado ver, uma disputa [acerca duma falha nas *Grundgesetze der Arithmetik*] apenas pode ocorrer no que diz respeito à minha lei básica de curso de valores (V). [...] De qualquer modo, marca-se aqui o lugar onde se tem de tomar uma decisão”. A decisão de Frege foi a de aceitar cursos de valores (e, portanto, extensões de conceitos). As extensões são objetos e nunca são confundidas nas *Grundgesetze der Arithmetik* com conceitos. Frege continua firmemente a submeter-se à terceira injunção. Porém, agora cada conceito tem um procurador no domínio dos objetos: a sua extensão. Como é sobejamente conhecido, isso é fatal. Depois da célebre carta de Russell a comunicar-lhe o paradoxo, Frege ainda foi a tempo de escrever um posfácio no segundo volume das *Grundgesetze der Arithmetik*. A seguinte passagem do posfácio é elucidativa: “Teria ficado feliz em ter dispensado esta fundamentação [extensões] se tivesse sabido dum substituto para ela. Mesmo agora, não vejo como é que a aritmética pode ser cientificamente fundamentada, como é que os números podem ser apreendidos como objetos lógicos e postos em consideração se não for permissível [...] passar dum conceito para a sua extensão”.

Desde o início do seu projeto de redução da aritmética à lógica, desde *Os Fundamentos da Aritmética*, Frege tentou evitar extensões. Na época d’*Os Fundamentos da Aritmética*, Frege ainda parecia alimentar a esperança de que as poderia dispensar. Finalmente, nas *Grundgesetze der Arithmetik*, resigna-se pragmaticamente a elas.

#### Notas e referências

Aproveitando a minha presença em 2017 no XVII Encontro Brasileiro de Lógica em Pirenópolis (estado de Goiás), o professor Nelson Gomes convidou-me a visitar Brasília e a dar um seminário na universidade. Quero aqui agradecer-lhe o convite e, em especial, o ter-me mostrado a cidade. Não podia ter tido um cicerone mais bem informado e mais agradável do que o professor Gomes. A 12 de maio fiz a referida comunicação ao Seminário Permanente “Lógica no Avião” da Universidade de Brasília. A palestra intitulou-se



---

“Ziguezague e a aritmética de Frege”. Na sequência desse seminário, o professor Alexandre Costa-Leite endereçou-me um convite para escrever um artigo. Porém, o material de que falei no seminário já estava comprometido para outra publicação, tendo entretanto saído à estampa em

Zigzag and Fregean arithmetic. In: “The Philosophers and Mathematics”, organizado por H. Tahiri. *Logic, Epistemology, and the Unity of Science* 43, Springer International, 2018, pp. 81-100.

Contudo, aceitei o convite do professor Costa-Leite e resolvi escrever o artigo que o leitor tem agora entre mãos. Debruço-me sobre os pontos do logicismo fregiano onde Frege é hesitante e menos claro. É, geralmente, nos pontos incertos do trabalho dum filósofo que recompensa um estudo mais atento. Muitas das observações que aqui faço são conhecidas e resisti a elaborar desenvolvimentos que surgiram naturalmente das considerações que discuto. O leitor interessado em mais informações e em referências bibliográficas pertinentes pode consultar a minha entrada “Logicismo” no *Compêndio em Linha de Problemas de Filosofia Analítica*, disponível em

<http://compendioemlinha.letras.ulisboa.pt/logicismo-fernando-ferreira/>

No presente artigo, fiz numerosas referências a dois trabalhos de Frege: ao *Die Grundlagen der Arithmetik* e aos dois tomos das *Grundgesetze der Arithmetik*. Para o primeiro trabalho, usei a seguinte tradução para o português:

Os Fundamentos da Aritmética. Tradução, prefácio e notas de António Zilhão. Imprensa Nacional-Casa da Moeda, Lisboa, 1992.

Para os dois tomos das *Grundgesetze der Arithmetik* usei a seguinte tradução para o inglês:

Basic Laws of Arithmetic. Tradução e edição de Philip A. Ebert e Marcus Rossberg. Oxford University Press, 2013.

As traduções para o português das *Grundgesetze der Arithmetik* são minhas, a partir do inglês acima. Listo, de seguida, os lugares onde o leitor pode encontrar as passagens citadas no texto:

- “mais do que ter tornado verosímil ...”, §90, Fundamentos.
- “o conteúdo de um juízo de reconhecimento”, §106, Fundamentos.
- “se a Inglaterra é o mesmo ...”, §66, Fundamentos.
- “Falta-nos o conceito ...”, §66, Fundamentos.
- “se este conhecido conquistador ...”, §56, Fundamentos.
- “pressupusemos como conhecido ...”, §107, Fundamentos.
- “Embora eu seja ...”, §68, Fundamentos.
- “como é que apreendemos ...”, posfácio, vol. 2, Basic Laws.
- “só seguindo este princípio ...”, §106, Fundamentos.
- “sem [cursos de valores] não ...”, prefácio, X, vol. 1, Basic Laws.
- “Agora, como é que se resolve ...”, §10, vol. 1, Basic Laws.
- “O matemático, tal como o geógrafo ...”, §96, Fundamentos.
- “pode ser apontado ...”, §146, vol. 2, Basic Laws.
- “Podemos chamar ...”, §147, vol. 2, Basic Laws.
- “Tanto quanto me é dado ver ...”, prefácio, VII, vol. 1, Basic Laws.
- “Teria ficado feliz ...”, posfácio, vol. 2, Basic Laws.

# SOME REMARKS ON THE VALIDITY OF THE PRINCIPLE OF EXPLOSION IN INTUITIONISTIC LOGIC<sup>1</sup>

Edgar Campos<sup>2</sup> & Abilio Rodrigues<sup>3</sup>

## Abstract

The formal system proposed by Heyting (1930, 1936) became the standard formulation of intuitionistic logic. The inference called *ex falso quodlibet*, or principle of explosion, according to which anything follows from a contradiction, holds in intuitionistic logic. However, it is not clear that explosion is in accordance with Brouwer’s views on the nature of mathematics and its relationship with logic. Indeed, van Atten (2009) argues that a formal system in line with Brouwer’s ideas should be a relevance logic. We agree that explosion should not hold in intuitionistic logic, but a relevance logic requires more than the invalidity of explosion. The principle known as *ex quodlibet verum*, according to which a valid formula follows from anything, should also be rejected by a relevantist. Given *ex quodlibet verum*, the inference we call weak explosion, according to which any negated proposition follows from a contradiction, is proved in a few steps. Although the same argument against explosion can be also applied against weak explosion, rejecting the latter requires the rejection of *ex quodlibet verum*. The result is the loss of at least one among reflexivity, monotonicity, and the deduction theorem in a Brouwerian intuitionistic logic, which seems to be an undesirable result.

## 1. Introduction: Brouwer on mathematics and logic

The main divergence between Brouwer and the standard approach to mathematics, based on classical logic, is his understanding of the notions of existence and truth as applied to mathematics. According to Brouwer, “to exist in mathematics means: to be constructed by intuition” (Brouwer, 1907, p. 96). So, the existence of mathematical objects such as numbers, functions, and sets depends on some construction introspectively carried out by some conscious subject.

The basic intuition of mathematics is the passage of time, which provides the raw material for the mathematical constructions and poses restrictions on which constructions may or may not be carried out (Brouwer, 1907, p. 52). The basic idea is that the human mind perceives the flow of time, the sequence of moments, one after another. A first moment is followed by a second, and this is something like a ‘pure form of time’ – indeed, an idea akin to Kantian pure intuition of time. The concept of a sequence of numbers is based on this basic intuition (Brouwer, 1912, pp. 127-128). The truth of any assertion about these mathematical objects depends on these constructions: “truth is only in reality i.e. in the present and past experiences of consciousness” (Brouwer, 1948, p. 1243).

---

<sup>1</sup> The first author acknowledges support from *CAPES* (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior), a doctorate scholarship. The second author acknowledges support from *CNPq* (Conselho Nacional de Desenvolvimento Científico e Tecnológico), research grant 311911/2018-8. The authors would like to thank Rodolfo Ertola-Biraben, André Porto and Wagner Sanz for some helpful comments and suggestions. They are not responsible, however, for any infelicities that may remain.

<sup>2</sup> Department of Philosophy, Federal University of Minas Gerais. Email: [ehncampos@gmail.com](mailto:ehncampos@gmail.com).

<sup>3</sup> Department of Philosophy, Federal University of Minas Gerais. Email: [abilio.rodrigues@gmail.com](mailto:abilio.rodrigues@gmail.com).

A consequence of such a view is his conception of logic being dependent upon mathematics. So conceived, logic just describes patterns of mental constructions in the language of mathematics, and the latter is a linguistic representation of these mental mathematical constructions. A valid formula in a Brouwerian intuitionistic logic represents something that can be constructed from the fundamental intuition of mathematics, and when a mathematician writes down a proof, such a proof is a linguistic *description* of those mental constructions.

Brouwer developed his well-known criticism of the principle of excluded middle, double negation elimination, and other classically valid principles, based on his views on the nature of mathematics and its relationship with logic. Excluded middle is not reliable because “it claims that every supposition is either true or false” (Brouwer, 1908, p. 109), which means, given that mathematical truth amounts to being mentally constructed, that every mathematical problem can be solved. The fact that there are open problems in mathematics implies that there is a proposition  $A$  such that no construction of  $A$ , nor of  $\neg A$  has been carried out; therefore, there is no guarantee that  $A \vee \neg A$  represents a mathematical construction.

In the standard formalization of intuitionistic logic (Heyting, 1930, 1956), the inference known as *ex falso quodlibet*, or the *principle of explosion*,

$$(EXP) A, \neg A \vdash B,$$

holds. The explanation of the meanings of the intuitionistic connectives, given by the *BHK*-interpretation, also takes *EXP* as valid (Troelstra and van Dalen, 1988). In the minimal logic proposed by Johansson (1937), although *EXP* does not hold, a weaker principle, that we call here *weak explosion*, holds:

$$(W-EXP) A, \neg A \vdash \neg B.$$

In both intuitionistic and minimal logic, and in classical logic as well, the principle known as *ex quodlibet verum*,

$$(EQV) \vdash A \rightarrow (B \rightarrow A),$$

holds. *EQV* is tantamount, in natural deduction systems, to the vacuous discharge of an assumption, and in sequent calculus, to the structural rule of weakening. The validity of *W-EXP* depends essentially on *EQV*.

The validity of *EXP* within the framework of a Brouwerian logic was criticized and rejected by Kolmogorov (1925) in his first work on intuitionistic logic, and also by Johansson (1937), where we find the formal system that became known as minimal logic. According to Haack (1974), Heyting’s justification of *EXP* in intuitionistic logic extends the intuitionistic sense of ‘construction’ in such a way that it could no longer be regarded as intuitionistic. As a result, she thinks that minimal logic is better to represent a Brouwerian intuitionistic logic than Heyting’s proposal (Haack, 1974, p. 102).

van Atten (2009) also defends the view that *EXP* should not be valid within a constructivist framework. However, van Atten goes further and argues that a formal system in line with Brouwer’s ideas should be a relevance logic. But if this were the case, not only *EXP* but also *EQV* and *W-EXP* should be considered invalid. Explosion, the deduction theorem and *EQV* yield the so-called paradoxes of material implication, and avoiding the latter is a motivation for relevance logics (see, for example, Mares, 2018). A relevantist, of course, also rejects the validity of *W-EXP*.

In this paper, we discuss the validity of *EXP*, *W-EXP*, and *EQV* in the framework of a Brouwerian intuitionistic logic. We agree with van Atten (2009) and Haack (1974) that explosion should not hold. On the other hand, although there may be some debate regarding the validity of *W-EXP* and *EQV*, both are much more plausible than explosion

in a constructive framework. We also suggest a normative justification for *W-EXP* in constructive mathematics as a way to prohibit contradictions. Although this idea does not fit with Brouwer's descriptive conception of logic, and cannot be found in his works, given the assumption that mathematics is consistent, *W-EXP* is more effective than the principle of non-contradiction in avoiding inconsistency.

The remainder of this text is structured as follows. In Section 2, we see Brouwer's remarks on the hypothetical judgment (implication) and the occurrence of contradictions. In Section 3, the passages in Kolmogorov (1925, 1932), Heyting (1930, 1934, 1956), Glivenko (1928, 1929) and Johansson (1937) where *EXP* is mentioned are presented, and we compare their views with Brouwer's remarks on implication and contradiction. In Section 4, we discuss how the *BHK*-interpretation justifies *EXP*. Section 4 also discusses the validity of disjunctive syllogism, *W-EXP*, and *EQV* in the framework of a Brouwerian intuitionistic logic.

## 2. The hypothetical judgment

The discussion about the validity of *EXP* depends on Brouwer's views on the implication. An implication  $A \rightarrow B$  is to be understood as a kind of 'path' that leads from  $A$  to  $B$ . So, whether or not *EXP* should be considered as valid, according to Brouwer's views depends on how any proposition  $B$  can be obtained from a pair of contradictory propositions  $A$  and  $\neg A$ . We divide the passage where Brouwer presents his views on implication and contradictions into two parts, that we call *I* and *II*.

### *I*

In one particular case the chain of syllogisms is of a somewhat different kind, which seems to come nearer to the usual logical figures and which actually seems to presuppose the hypothetical judgment from logic. This occurs when a structure is defined by some relation in another structure, while it is not immediately clear how to effect its construction. Here it seems that the construction is supposed to be effected, and that starting from this hypothesis a chain of hypothetical judgments is deduced. But this is no more than apparent; what actually happens is the following: one starts by setting up a structure which fulfills part of the required relations, thereupon one tries to deduce from these relations, by means of tautologies, other relations, in such a way that these new relations, combined with those that have not yet been used, yield a system of conditions, suitable as a starting-point for the construction of the required structure. Only by this construction will it be proved that the original conditions can be fulfilled. (Brouwer, 1907, p. 72-73)

By a 'hypothetical judgment from logic' Brouwer means a sequence of one or more inferences that start from an assumption  $A$  and reach a conclusion  $B$ . The result of such a reasoning is an implication  $A \rightarrow B$ . Brouwer adds:

### *II*

'But', the logician will retort, 'it might have happened that in the course of these reasonings a contradiction turned up between the newly deduced relations and those that had been kept in store. This contradiction, to be sure, will be observed as a logical figure, and this observation will be based upon the principium contradictionis.' To this we can reply: 'The words of your mathematical demonstration merely accompany a mathematical construction that is effected without words. At the point where

you enounce the contradiction, I simply perceive that the construction no longer goes, that the required structure cannot be imbedded in the given basic structure. And when I make this observation, I do not think of a principium contradictionis. (Ibidem)

In *I*, Brouwer talks about implication, and in *II*, about the principle of noncontradiction. These passages, especially *I*, yield some interpretation issues. We start with the easiest one, quotation *II*.

**2.1. The principle of non-contradiction.** We want to draw your attention to two points in quotation *II*. The first is that a contradiction, as expected, given the assumption that mathematics is consistent, cannot be constructed. So,

the principle of contradiction is indisputable: The results that we perform the imbedding of a system  $a$  into a system  $b$  in a prescribed manner, and that we are arrested by the impossibility of such an imbedding, exclude each other (Brouwer, 1908, p. 109).

A contradiction is the sign that the intended construction cannot go further. Hence, the main role for a contradiction is to show that a hypothetical reasoning has reached a dead end and has to stop. Note that a contradiction can only appear from a reasoning that starts from premises, that is, a hypothetical reasoning. This is the basis of the intuitionistic interpretation of negation defined in terms of the primitive notion of *absurd*,  $A \rightarrow \perp$ .

The second is the claim that logic is secondary to the mental constructions carried out by the mathematician. It is not that a construction cannot be done because it disobeys non-contradiction. Quite the contrary, non-contradiction holds *because* no construction of a contradiction can be done. The principle of non-contradiction  $\neg(A \wedge \neg A)$ , and the notion of absurd, represented by  $\perp$ , are just linguistic representations of mental mathematical constructions that cannot go further.

**2.2. Implication.** Now, we turn to quotation *I*, where Brouwer presents his views on implication. van Dalen (2004) proposed the following reading:

Brouwer’s comments on hypothetical judgments are unfortunately rather cryptic. They can be read in various ways. The most plausible reading seems to be the following.

( $\alpha$ ) In order to establish  $A \rightarrow B$ , one has to carry out two tasks, namely, (i) find a construction for (the structure specified by)  $A$ , (ii) find a construction for (the structure specified by)  $B$  that departs from the first construction. We have left the reference to the “embedding” implicit. In fact, this embedding is mostly tacitly incorporated into the construction for  $B$  (van Dalen, 2004, p. 251).

Although van Dalen says in the quotation above that this reading is the ‘most plausible’, he will later reject this (van Dalen, 2008, p. 20). His interpretation was indeed criticized by van Atten (2009), who proposed the following interpretation:

( $\beta$ ) In order to establish  $A \rightarrow B$ , one has to conceive of  $A$  and  $B$  as conditions on constructions, and to show that from the conditions specified by  $A$  one obtains the conditions specified by  $B$ , according to transformations whose composition preserves mathematical constructibility (van Atten, 2009, p. 128).

As van Atten points out, the central aspect of interpretation  $\beta$  is the notion of *condition*: “Brouwer avoids commitment to hypothetical constructions by considering, not constructions, but conditions on constructions”, and such conditions are “actual and therefore ontologically unproblematic” (van Atten, 2009, p. 127).

According to  $\alpha$ , an implication  $A \rightarrow B$  implies that both  $A$  and  $B$  have been effectively constructed: first, a construction of  $A$  has to be already done, and so  $B$  is constructed departing from  $A$ . Thus,  $\alpha$  excludes the possibility of putting forward  $A$  as a mere hypothesis. Notice that this reading makes it impossible even to prove  $\neg A$ , since such a proof would need an effective construction of  $A$ , which is precisely what cannot be done when  $\neg A$  holds.

The interpretation  $\beta$ , on the other hand, understands  $A$  as a condition on a construction, and so, unlike  $\alpha$ , does not require that  $A$  has been effectively constructed. So, a condition for a construction can be, so to speak, just ‘imagined’, and what can be done from these conditions is obtained by means of steps that go from one condition to another condition. Thus, an implication  $A \rightarrow B$  can be read as ‘if the condition  $A$  is satisfied, then the condition  $B$  is satisfied’, and if eventually it is proved that condition  $B$  cannot be satisfied, i.e.,  $\neg B$  is constructed, we can draw the conclusion that condition  $A$  cannot be satisfied either, so  $\neg A$ . These reasonings are clearly constructive. This interpretation also acquires support from a claim made by Brouwer in one of his last publications, where he talks about a ‘constructional condition’ on a mathematical system:

the wording of a mathematical theorem has no sense unless it indicates the construction either of an actual mathematical entity or of an incompatibility (e.g., the identity of the empty two-ity with an empty unity) out of some constructional condition imposed on a hypothetical mathematical system (Brouwer, 1954, p. 3).

The rejection of *EXP* by interpretation  $\alpha$  is straightforward. If an implication  $A \rightarrow B$ , in order to be a legitimate mathematical construction, requires that a construction of the antecedent  $A$  has been successfully carried out, and since a construction of a contradiction is never carried out, clearly nothing can be constructed from a contradiction. The interpretation  $\beta$  also rejects the validity of *EXP*. Indeed, it is not clear how to understand a contradiction as a condition of a construction that, in the end, would obtain an arbitrary proposition  $B$ , or better, an arbitrary condition expressed by a proposition  $B$ . Therefore, both interpretations of Brouwer’s remarks on the hypothetical judgment provide reasons for rejecting the validity of *EXP* in a Brouwerian intuitionistic logic.<sup>4</sup>

### 3. Kolmogorov, Heyting, Glivenko, Johansson

As we have seen, Brouwer has not provided any reasons for accepting *EXP* in his works, actually, quite the contrary, we can find strong evidence for rejecting it. However, *EXP* holds in the formalization of intuitionistic logic that became standard. Different formal systems, conceived to express Brouwer’s ideas, were proposed by Kolmogorov (1925), Glivenko (1928, 1929), and Heyting (1930, 1956). In Johansson (1937) a formalization of intuitionistic logic was proposed, although he did not explicitly mention Brouwer’s work.

**3.1. Kolmogorov 1925.** The first formal system designed to express Brouwer’s ideas was proposed by Kolmogorov (1925), in a language with only negation and implication.

---

<sup>4</sup> van Atten (2009, p. 130) reaches a similar conclusion: “we cannot expect to be able to come up with the required transformation for just any given pair of a false antecedent and a consequent, let alone that we have a uniform method to do this. As a consequence, there is on this account of the hypothetical judgement no ground to accept the Ex Falso principle [*EXP*] in its full generality.”

He proposed a propositional logic by dropping the two axioms of negation presented in Hilbert (1923),

$$A \rightarrow (\neg A \rightarrow B),$$

and

$$(A \rightarrow B) \rightarrow ((\neg A \rightarrow B) \rightarrow B)$$

and adding only one axiom for negation, that we call here *introduction of negation*,

$$(IN) (A \rightarrow B) \rightarrow ((A \rightarrow \neg B) \rightarrow \neg A),$$

that correctly expresses the notion of intuitionistic negation (Section 2.1 above). In Kolmogorov's system, *EQV* holds and *W-EXP* can be proved. About the rejection of *EXP*, Kolmogorov remarks that:

Hilbert's first axiom of negation, 'Anything follows from the false' (...) does not have and cannot have any intuitive foundation since it asserts something about the consequences of something impossible: we have to accept *B* if the true judgment *A* is regarded as false (Kolmogorov, 1925, p. 421).

Thus, in 1925, Kolmogorov defended the view that it makes no sense to speak about what would be constructed having as a starting point something that cannot be constructed. Notice that this claim is in line with the argument for the rejection of *EXP* based on Brouwer's views, discussed in Section 2 above. Later, however, Kolmogorov changed his mind. In 1932, when he presents the interpretation of propositional connectives in terms of solutions of problems, he says that

*the calculus of problems is formally identical with the Brouwerian intuitionistic logic, which has recently been formalized by Mr. Heyting* (Kolmogorov, 1932, p. 328).

He refers to Heyting (1930), where a formal system for intuitionistic logic, in which *EXP* holds, has been presented (Section 3.2 below). Note that there is no evidence that Kolmogorov's article published in 1925 was known by Heyting before the latter conceived his formal system. Regarding the justification of the validity of *EXP*, now we read "As soon as  $\neg a$  is solved, then the solution of *a* is impossible and the problem  $a \rightarrow b$  is without content" (Kolmogorov, 1932, p. 331). This interpretation intends to validate the principle of explosion because it presupposes that "the proof that a problem is without content will always be considered as its solution" (Kolmogorov, 1932, p. 329). So, once  $\neg a$  has been solved,  $a \rightarrow b$  is also considered solved.<sup>5</sup> This interpretation is not implausible, but we do not think that it fits Brouwer's constructivist ideas.

**3.2. Heyting 1930.** Heyting presented his version of intuitionistic logic during a prize contest promoted by the Dutch Mathematical Society in 1927. Although the original manuscript has not survived, a revised version was published in 1930. This system, presented in a language with  $\wedge$ ,  $\vee$ ,  $\rightarrow$ , and  $\neg$ , contains the axiom

$$4.1 \neg a \rightarrow (a \rightarrow b) \text{ (Heyting, 1930, p. 313).}$$

Heyting's system became the standard intuitionistic logic, either with  $\neg$  as primitive, or with  $\perp$  as primitive and negation  $\neg A$  defined as  $A \rightarrow \perp$ .

This work was favorably received by Brouwer, as it is clear from his letter to Heyting, dated July 17th, 1928 (van Dalen, 2011, p. 333-334). We do not think, however, that such positive reception of Heyting's paper by Brouwer should be taken as evidence that the latter would accept the validity of *EXP*. Supporting a publication does not imply

<sup>5</sup> We use here capital letters as propositional variables. However, we keep the notation used by the authors in the quotations, and in our comments related to such quotations.

the complete agreement with its contents – indeed, in the 1940s, Brouwer supported Griss’ publications on negationless mathematics, even though he managed to publish a counterexample to Griss’ ideas.

Heyting provides an informal explanation for the meaning of the implication that intends to justify both *EQV* and *EXP*:

The case is conceivable that after the statement  $a \rightarrow b$  has been proved in the sense specified, it turns out that  $b$  is always correct. Once accepted, the formula  $a \rightarrow b$  then has to remain correct; that is, we must attribute a meaning to the sign  $\rightarrow$  such that  $a \rightarrow b$  still holds. The same can be remarked in the case where it later turns out that  $a$  is always false (Heyting, 1930, p. 313).

Heyting’s argument, however, as Johansson will point out in 1937 (Section 3.4 below), is not sound. He was thinking of a situation in which the proof of  $A \rightarrow B$  was available, and subsequently a proof of  $\neg A$  was obtained. But what *EXP* says is something rather different, and stronger, namely, that a proof of  $\neg A$  automatically yields a proof of  $A \rightarrow B$ . Heyting clearly fails to justify this.

**3.3. Glivenko 1928.** In 1928, before the revised version of his 1927 paper was published by Heyting, Glivenko (1928), motivated by criticisms directed to Brouwer’s intuitionism that he thought were mistaken, proposed a list of formulas that should be valid according to Brouwer’s views. In this list, nothing equivalent to *EXP* appears. But in the following year, Glivenko added more axioms to the previous list of formulas of 1928 in order to “lay the complete foundations of a logical calculus” (Glivenko, 1929, p. 302). Among these axioms, we find (C) and (D):

- (C)  $p \rightarrow (q \rightarrow p)$ ,  
 (D)  $\neg q \rightarrow (q \rightarrow p)$ .

He then asks about the validity of these axioms and provides the following explanation:

Are these axioms C and D admissible in the Brouwerian logic? I will confine myself here to a simple remark. By virtue of axioms VI and VII [resp.  $p \rightarrow (p \vee q)$  and  $q \rightarrow (p \vee q)$ ], expressions C and D are nothing more than the immediate consequences of the principle:

$$(p \vee \neg q) \rightarrow (q \rightarrow p)$$

whose admissibility is quite evident, because the formal implication  $p \rightarrow q$  has no other sense than “when one accepts the truth of  $p$ , one must accept that of  $q$ ” (Glivenko, 1929, p. 302).<sup>6</sup>

The principle  $(p \vee \neg q) \rightarrow (q \rightarrow p)$  is equivalent, given the deduction theorem, to the inference known as the disjunctive syllogism. Indeed, if disjunctive syllogism holds, *EXP* and *EQV* are easily proved. However, it is far from clear that disjunctive syllogism should be valid in intuitionistic logic – actually, we argue in Section 4.1 below that it should not be valid.

---

<sup>6</sup> In a footnote (Ibidem pp. 304-305), Glivenko remarks that it was Heyting who made him acknowledge the appropriateness of *EXP* in the Brouwerian logic and refers the reader to Heyting’s revised version of his formalization that was going to be published in the *Mathematische Annalen*. However, because of Brouwer’s dismissal of the Journal’s editorial board by Hilbert in 1928, the episode that became known as the *Grundlagenstreit*, Heyting’s article was published elsewhere.



**3.4. Johansson 1937.** In Johansson (1937) a formal system that became known as *minimal logic* was presented, in a language with  $\vee$ ,  $\wedge$ ,  $\rightarrow$ , and  $\neg$ , whose only axiom for negation is the introduction of negation, expressed as

$$4.11 ((a \rightarrow b) \wedge (a \rightarrow \neg b)) \rightarrow \neg a \text{ (Johansson, 1937, p. 121).}$$

Note that this axiom is equivalent, in a language with  $\wedge$ , to the axiom *IN* proposed by Kolmogorov (1925) (see Section 3.1 above). Indeed, it is well-known that Kolmogorov's system is the fragment of minimal logic in a language with  $\rightarrow$  and  $\neg$ . Johansson opens his article by mentioning two formulas accepted by Heyting (1930),

$$2.14 b \rightarrow (a \rightarrow b),$$

and

$$4.1 \neg a \rightarrow (a \rightarrow b),$$

and casts doubt on the validity of them. The first is *EQV*, the second is tantamount to *EXP*. Regarding 4.1, he says that the case in which one writes  $a \rightarrow b$  when  $a$  is false or absurd,

means a difficult-to-overlook extension of the concept of [logical] consequence. It will be worth the effort to investigate if one cannot avoid it (Johansson, 1937, p. 119, our translation).

He ends up accepting 2.14 (*EQV*) but rejecting 4.1 (and so *EXP*) in his minimal logic. Johansson did not justify in this paper the rejection of *EXP*, but he expressed his concerns about *EXP* in a letter to Heyting dated September 23rd, 1935, prior to the 1937 article:

I have not been able to make my peace with Axiom 4.1. You say that when  $a \rightarrow b$  has been proved, and later  $\neg a$  is proved, then  $a \rightarrow b$  should remain correct. Indeed; but  $\neg a \rightarrow (a \rightarrow b)$  means that when  $\neg a$  has been proved,  $b$  at once becomes derivable from  $a$ , even when this had not been proved before. And that contradicts my intuition in the most violent way (Letter from Johansson to Heyting, *apud* van Atten, 2017).

We think that Johansson's argument above succeeds in refuting Heyting's justification of *EXP* (Section 3.2 above).

**3.5. Heyting 1956.** In 1955, a year before the publication of his *Intuitionism: an Introduction*, Heyting adds a short note to the French translation of Heyting (1934):

JOHANSSON [1] developed a 'minimal calculus' in which he adopts all the axioms from HEYTING except  $\neg a \rightarrow (a \rightarrow b)$ : one may consider this calculus as the expression of the intuitionistic logic with another interpretation of the implication (Heyting, 1955, p. 19, our translation).

As van Atten (2017, Sec. 6.1) suggests, the conversation with Johansson may have had an effect on Heyting's position regarding *EXP*. In Heyting (1956), an effort is made to justify *EXP* along Brouwerian lines as much as possible:

Axiom X [ $\neg p \rightarrow (p \rightarrow q)$ ] may not seem intuitively clear. As a matter of fact, it adds to the precision of the definition of implication. You remember that  $p \rightarrow q$  can be asserted if and only if we possess a construction which, joined to the construction  $p$ , would prove  $q$ . Now suppose that  $\neg p$ , that is, we have deduced a contradiction from the supposition that  $p$  were carried out. Then, in a sense, this can be considered as a construction, which, joined to a proof of  $p$  (which cannot exist) leads to a proof of  $q$ . I shall interpret the implication in this wider sense (Heyting, 1956, p. 106).

It seems that Heyting wants to justify *EXP* along the same lines as he did for the other axioms of intuitionistic logic, namely, by presenting a construction that allows one to obtain a proof of an arbitrary proposition  $q$  from any proof of  $p$ , once  $\neg p$  has been constructed. But one can hardly take his attempt as successful. It is difficult to understand, *in Brouwerian terms*, how a construction of  $\neg p$  together with a non-existent construction of  $p$  could yield a construction of an arbitrary  $q$ .

#### 4. Explosion in the *BHK*-interpretation

After Heyting's last effort, we find other attempts at the justification of *EXP* in intuitionistic logic. One comes from the standard proof interpretation for the logical constants, developed from the contributions of Brouwer, Heyting and Kolmogorov, known as the *BHK*-interpretation. The origins of the *BHK*-interpretation can be traced back to Kolmogorov's proposal of interpreting intuitionistic logic as being concerned with solutions of problems. Heyting (1956, p. 102) explains the intuitionistic connectives in terms of assertibility conditions. Later, Troelstra and van Dalen (1988) presented what became the standard *BHK*-interpretation in terms of proofs. They explain the validity of *EXP* as follows:

$\perp \rightarrow A$  is generally provable: since there is no proof of  $\perp$ ,  $\lambda a.a$  (or any other mapping) may count as a proof of  $\perp \rightarrow A$ , since it has to be applied to an empty domain. The principle  $\perp \rightarrow A$  ("ex falso sequitur quodlibet") has sometimes been rejected as non-constructive (Johansson, 1937); (Heyting, 1956, 7.1.3) regards it as an extra stipulation fixing the meaning and use of  $\perp$ ,  $\rightarrow$  Troelstra and van Dalen (1988, p. 10).

Notice that van Dalen and Troelstra consider Heyting's explanation (quoted above in Section 3.5) as a *stipulation* that fixes the meaning of  $\perp$  and  $\rightarrow$ , or  $\neg$  and  $\rightarrow$ , depending on what one chooses as primitive. But if this inference were in accordance with Brouwer's ideas, it would not have to be the result of a stipulation. As far as we know, nowhere in Brouwer is there evidence in support of the claim that any operation (i.e., a construction), applied to an empty domain (i.e., to a non-existent proof of  $\perp$ ) yields as a result a proof of an arbitrary proposition  $A$ .

Moreover, we think that this explanation has a circular ingredient. The argument that goes from 'there is no proof of  $\perp$ ' to 'any given proof of  $\perp$  can be transformed in a proof of  $A$ ' seems to assume *EXP* in the meta-theory. There is a striking similarity between the claim that any operation applied to nothing yields anything and, for example, the proof that the empty set is a subset of any set, which uses explosion. Given that  $\forall x \neg(x \in \emptyset)$ , it follows that  $\forall x \forall A(x \in \emptyset \rightarrow x \in A)$ . It holds precisely because the antecedent  $x \in \emptyset$  is false for any  $x$ . Classical logic, of course, explicitly relies on a classical meta-theory for the explanation of all the connectives. But it is doubtful that a Brouwerian intuitionistic logic should do the same regarding an inference that has no clear justification when we consider Brouwer's views.

**4.1. Disjunctive syllogism.** We saw, in Section 3.3 that Glivenko (1929) justifies the intuitionistic validity of *EXP* based on the fact that it follows from disjunctive syllogism:

$$(DS) A \vee B, \neg A \vdash B.$$

Indeed, there is a well-known proof of the principle of explosion that can be traced back at least to medieval times:

$$\frac{\frac{A}{A \vee B} \vee I}{B} \neg A \text{ } DS$$

This proof makes essential use of disjunctive syllogism (*DS*). van Atten (2009) argues that although *EXP* should be invalid within a constructive framework, there is no problem with disjunctive syllogism in a Brouwerian logic. He argues that the inference rules applied in the argument are indeed correct from a Brouwerian point of view, but the whole argument, from  $A$  and  $\neg A$  to  $B$ , is unreliable because there is no guarantee that the *composition* of the inference rules preserves constructibility (van Atten, 2009, p. 124). In other words, constructibility is not transitive, which means that a logic suitable for constructive mathematics cannot be transitive.

It is well-known that given a few assumptions *EXP* and *DS* are equivalent. Indeed, given *DS*, *EXP* can be proved with introduction of disjunction (proof above), and given *EXP*, *DS* can be proved with elimination of disjunction:

$$\frac{A \vee B \quad \frac{[A]^1 \quad \neg A}{B} \text{EXP} \quad [B]^1}{B} \vee E,1$$

Now, since the validity of both  $\vee I$  and  $\vee E$  do not conflict with Brouwer's ideas, the rejection of *EXP* should imply the rejection of *DS*. The constructive validity of *DS* has also been discussed by Tennant (1987). Both van Atten (2009) and Tennant rejected *EXP* and accepted *DS* as valid, but the price paid is the loss of transitivity. We think, however, that there is no reason to reject transitivity in the framework of mathematics, including constructive mathematics. Brouwer agrees with this idea, as the following passage clearly shows:

[the inference] that concludes from the imbedding of a system  $b$  into a system  $c$ , joined to the imbedding of a system  $a$  into the system  $b$ , to a direct imbedding of the system  $a$  into the system  $c$  ... is nothing more than a tautology (Brouwer, 1908, p. 109).

By being 'nothing more than a tautology' Brouwer means that this inference, namely, transitivity, is constructively valid.

In intuitionistic logic, a disjunction  $A \vee B$  holds if and only if  $A$  has been constructed or  $B$  has been constructed. But since  $\neg A$  (the minor premise of *DS*) has been constructed, a construction of  $A$  is not possible. Therefore, it has to be the case that the disjunction  $A \vee B$  was the result of  $\vee I$  from the premise  $B$ , and not from premise  $A$ . This is the same line of reasoning as found in van Atten (2009, p. 124). Nevertheless, rather than showing that *DS* is valid, in our view, it shows that *DS* is not necessary at all. So, once the validity of *EXP* is rejected, the proof of *EXP* from *DS*, mentioned at the beginning of this section, is not possible. Rather, it should be replaced by a derivation of  $B$  that was the only alternative to obtain constructively  $A \vee B$ , given the impossibility of constructing  $A$ .

**4.2. Ex quodlibet verum.** Now, we turn to *EQV*. Is the validity of the axiom 2.14 of Heyting (1956) indisputable in intuitionistic logic? As a matter of fact, we find no criticisms of *EQV* in the intuitionistic tradition. However, Johansson (1937) rejected the validity of explosion in intuitionistic logic because it seemed wrong to him to claim that when  $\neg A$  has been proved, any  $B$  at once becomes derivable from  $A$ , even when this had not been proved before (see Section 3.4). A similar but weaker argument could be used to cast doubt on *EQV*, since it says that when  $B$  has been proved,  $B$  at once becomes derivable from any  $A$ , even when this had not been proved before.

The proof of *EQV* in a natural deduction system is as follows:

$$\frac{\frac{[A]^1}{B \rightarrow A} \rightarrow I}{A \rightarrow (B \rightarrow A)} \rightarrow I,1$$

The second application of  $\rightarrow I$ , that concludes *EQV*, discharges the assumption  $A$ . But the first application of  $\rightarrow I$  discharges an assumption that has not been declared. The introduction rule for implication allows the discharge of a vacuous assumption, and this move is tantamount to *EQV* in an axiomatic system. The rationale of this is that if a construction of  $A$  is already given, any construction  $B$ , so to speak, can be ‘placed on the top’ of that construction of  $A$ :

$$\begin{array}{ccc} & & \vdots \\ & & B \\ & & \vdots \\ \vdots & \text{hence} & A \\ A, & & \vdots \end{array}$$

So, the construction of  $A$  ‘follows’ from  $B$  just because it was already there. We think, however, that this argument can be improved with the help of van Atten’s interpretation  $\beta$ . First, note that reflexivity holds according to  $\beta$ . Given a condition  $A$ , the same condition  $A$  is obtained, so,  $A$  leads to  $A$ :

$$A \vdash A.$$

We cannot see anything non-constructive in this reasoning. Now, given a condition  $A$ , any other condition  $B$  added to  $A$  still yields condition  $A$ , since anything that is obtained from  $A$  can also be obtained from any  $B$  that is put together with  $A$ . This is nothing but monotonicity, and there is no reason to suppose that constructive mathematical reasoning is non-monotonic. So, given reflexivity and monotonicity,

$$A, B \vdash A$$

holds, and two applications of the deduction theorem<sup>7</sup> yield

$$\vdash A \rightarrow (B \rightarrow A).$$

Again, we cannot see anything non-constructive in this line of reasoning.

**4.3. The weak explosion.** As we have seen, in the standard formalization of intuitionistic logic, the principle of explosion holds. In the formal systems proposed by Kolmogorov (1925) and Johansson (1937) (minimal logic), although *EXP* does not hold, a weaker principle, called here *weak explosion*, holds:

$$(W-EXP) A, \neg A \vdash \neg B.$$

It may seem that both *EXP* and *W-EXP* should be invalid in a Brouwerian intuitionistic logic because *W-EXP* is vulnerable to the same argument that justified the rejection of *EXP* in Section 2. Indeed, if there cannot be a constructive method that transforms the conditions on the construction of a contradiction into the conditions on the construction of an arbitrary proposition  $B$ , the same reasoning applies when  $B$  is  $\neg B$ . However, this point is not so simple. The proof of *W-EXP* depends essentially on *EQV*, which holds in intuitionistic logic and minimal logic. The proof of *W-EXP* is straightforward: from a pair of contradictory propositions  $A$  and  $\neg A$ , from *EQV*, we get  $B \rightarrow A$  and  $B \rightarrow \neg A$ , and so, applying the introduction of negation and modus ponens,  $\neg B$  is obtained. Similarly, if we take  $\perp$  as primitive, from *EQV* we get

$$\vdash \perp \rightarrow (B \rightarrow \perp),$$

<sup>7</sup> It is not clear whether or not Kolmogorov (1925), Heyting (1930), Glivenko (1929) and Johansson (1937) had the deduction theorem – actually, it is likely that they didn’t. However, the deduction theorem holds in all these formal systems.

and so,

$$\perp \vdash B \rightarrow \perp.$$

What we have here is a sort of an aporetic situation: on one hand, a plausible argument that concludes the invalidity of *EXP* and can be extended to *W-EXP*; on the other hand, an even more plausible justification of *EQV*.

**4.4. On the normative character of *EXP* and *W-EXP*.** Although *W-EXP* does not make triviality tantamount to contradictoriness, a scenario in which all negated propositions can be inferred from any contradiction is not really any better than triviality. A logic in which *EXP* holds is called explosive, and a logic such that *EXP* does not hold but *W-EXP* holds is called partially explosive (Carnielli, Coniglio, and Marcos, 2007, p. 14).

Paraconsistent logics are formal systems in which the principle of explosion does not hold. A theory whose underlying logic is paraconsistent can accept contradictions without falling into triviality. It is a fact that several contexts of reasoning, from databases to scientific theories, are contradictory or yield contradictions in some critical situations.<sup>8</sup> The motivation for paraconsistent logics is to have a formal system capable of dealing with such contexts in which contradictions are expected to occur. Clearly, paraconsistent logics cannot be partially explosive, and *W-EXP* cannot be valid. A paraconsistent logic in which *W-EXP* does not hold is called *boldly paraconsistent* (Carnielli, Coniglio, and Marcos, 2007, p. 14).

If *EXP* does not hold in a Brouwerian intuitionistic logic, the latter turns out to be a paraconsistent logic. However, this is not of much significance. Mathematicians, and Brouwer as well, work based on the assumption that mathematics is free of contradictions. The invalidity of *EXP* in a logic designed to express Brouwer's ideas is not because contradictions are possible, or eventually might appear. Indeed, Brouwer's remarks about non-contradiction (see Section 2.1 above) make it clear that a contradiction cannot be constructed. Besides, as we saw in Section 4.3 above, the argument for rejecting *EXP* in the intuitionistic framework is convincing, and a similar argument may be put forward against *W-EXP*. However, the rejection of *W-EXP* requires the rejection of *EQV*, and the latter, in turn, as we have just seen, is a consequence of the deduction theorem, reflexivity and monotonicity.

The validity of *W-EXP* in intuitionistic logic may be explained from a normative point of view. Let us take a look at the role of *EXP* in classical logic. A central point of logic, no matter whether it is classical, intuitionistic or paraconsistent, is its normative character: the logical principles tell us how to reason correctly. In the case of intuitionistic logic, this normativeness comes not directly from logic, since the latter is a description of correct mental constructions. But once these principles are established, there is nothing wrong in taking them as normative. The classical notion of logical consequence is defined in terms of preservation of truth (of course, a transcendent notion of truth that is rejected by Brouwer). Hence, if one wants to reason correctly, in order to avoid obtaining false conclusion from true premises, one has to reason according to classical logic. The normative character of paraconsistent logic is clear: in a contradictory context of reasoning, *EXP* must not be applied, because although contradictions may occur, and do occur, triviality is unacceptable.

<sup>8</sup> See, for example, Belnap (1977) and Nickles (2002). Adopting a paraconsistent logic, however, does not imply the dialetheist view according to which there are true contradictions (e.g. Priest and Berto, 2013). The contradictions accepted by paraconsistent logics can be interpreted as conflicting evidence (Carnielli and Rodrigues, 2017) or inconsistent information (Belnap, 1977). In both cases, a property weaker than truth is attributed to contradictory propositions, in the sense that a proposition may enjoy that property without being true.

The validity of *EXP* in classical logic is the result of a decision that took preservation of truth as both a necessary and sufficient condition for logical consequence. So, the justification of *EXP* in classical logic is as follows: preservation of truth is a sufficient condition for logical consequence, and there is no model  $M$  such that both  $A$  and  $\neg A$  are true in  $M$ ; therefore, for any  $B$ , there is no model  $M$  such that  $A$  and  $\neg A$  are true in  $M$  and  $B$  is false in  $M$ . Although this justifies the validity of *EXP*, the role of *EXP* in classical logic has not been explained yet. Of course, *EXP* is never *used* in classical reasoning, except for proving some other inferences – the disjunctive syllogism for example – that nevertheless already have an intuitive appeal independent from *EXP*. In our view, what explains the role of *EXP* in classical logic is its normative character. The principle of non-contradiction prohibits contradictions by saying that  $A \wedge \neg A$  is always false. Explosion, in its turn, says something stronger: it explicitly states the consequences of accepting a contradiction. It is like two notices, one saying ‘smoking not allowed here’ and other saying ‘smoking not allowed here – penalty: £200’. Although from a strictly Kantian point of view, the consequences of the act should not interfere in the moral action, in fact – unfortunately maybe – the second notice is likely to be much more effective in avoiding smoking than the first. Something analogous occurs with respect to the principles of non-contradiction and explosion. The former just says that one cannot accept a contradiction because a contradiction is always false, while the latter says the consequences one is committed to if she/he accepts a contradiction. So, the role of *EXP* in classical logic is rather normative. It does not describe any real situation precisely because such a situation is not possible. Clearly, from the assumption that mathematics is free of contradictions, a pair of contradictory propositions can never be proved, either classically, or intuitionistically. So, it is reasonable to understand, and to explain, the validity of both *EXP* and *W-EXP* in a constructive logic designed to be a logic of constructive mathematical reasoning as a way of explicitly prohibiting contradictions.

## 5. Final remarks

In this paper, we discussed, from the point of view of Brouwer’s conception of constructive mathematics, three principles that hold in the standard formulation of intuitionistic logic, namely, *EXP* (explosion, or *ex falso quodlibet*), *W-EXP* (weak explosion) and *EQV* (*ex quodlibet verum*). There are plausible reasons for rejecting *EXP*, and on this point we agree with van Atten (2009). But van Atten also defends the validity of disjunctive syllogism in a Brouwerian logic and suggests that the latter should be a relevance logic. In both cases, we do not agree with him. A relevance logic should reject *EXP* and *EQV*. The same argument against *EXP* can be directed against *W-EXP*. But we saw in Section 4.3 that in order to reject *W-EXP*, since the latter follows from *EQV*, at least one from reflexivity, monotonicity and the deduction theorem has to be considered invalid. In Section 4.2, we argued that these inferences should hold in a Brouwerian intuitionistic logic. Regarding disjunctive syllogism, van Atten’s argument against it implies the loss of transitivity. In Section 4.1, we have shown that there is evidence in Brouwer’s works that transitivity holds in constructive mathematics. We suggested also, in Section 4.4, that the validity of *W-EXP* in minimal logic may be understood with a normative role, even though this idea does not fit the descriptive character of logic defended by Brouwer.

Our conclusion here, however, is that although there are strong reasons for rejecting *EXP*, whether or not *EQV* and *W-EXP* should also be rejected in a Brouwerian intuitionistic logic is an issue not settled yet. More detailed investigations into Brouwer’s work and into constructive results in mathematics are needed.

## Bibliography

- N. Belnap. How a computer should think. In *Contemporary Aspects of Philosophy* (ed. G. Ryle). Oriel Press, 1977.
- L. E. J. Brouwer. On the foundations of mathematics. In *Collected Works vol. I.* (ed. A. Heyting). North-Holland Publishing Company (1975), 1907.
- L. E. J. Brouwer. The unreliability of the logical principles. In *Collected Works vol. I.* (ed. A. Heyting). North-Holland Publishing Company (1975), 1908.
- L. E. J. Brouwer. Intuitionism and formalism. In *Collected Works vol. I.* (ed. A. Heyting). North-Holland Publishing Company (1975), 1912.
- L. E. J. Brouwer. Consciousness, philosophy and mathematics. In *Collected Works vol. I.* (ed. A. Heyting). North-Holland Publishing Company (1975), 1948.
- L. E. J. Brouwer. Points and spaces. In *Collected Works vol. I.* (ed. A. Heyting). North-Holland Publishing (1975), 1954.
- W. Carnielli and A. Rodrigues. An epistemic approach to paraconsistency: a logic of evidence and truth. *Synthese*, 2017. Preprint <http://philsci-archive.pitt.edu/14115/1/letj.pdf>.
- W. Carnielli, M. E. Coniglio, and J. Marcos. Logics of formal inconsistency. In *Handbook of Philosophical Logic* (eds. Dov Gabbay and Franz Guenther), volume 14. Springer, 2007.
- V. Glivenko. Sur la logique de M. Brouwer. *Académie Royale de Belgique. Bulletin de la Classe des Sciences*, 14:225–228, 1928.
- V. Glivenko. On some points of the logic of Mr. Brouwer. In *From Brouwer To Hilbert: The Debate on the Foundations of Mathematics in the 1920s* (ed. P. Mancosu). Oxford University Press (1998), 1929.
- S. Haack. *Deviant Logic: Some Philosophical Issues*. Cambridge University Press, 1974.
- A. Heyting. The formal rules of intuitionistic logic. In *From Brouwer To Hilbert: The Debate on the Foundations of Mathematics in the 1920s* (ed. P. Mancosu). Oxford University Press (1998), 1930.
- A. Heyting. *Mathematische Grundlagenforschung, Intuitionismus, Beweistheorie*. Springer-Verlag, 1934.
- A. Heyting. *Les Fondements des Mathématiques, Intuitionnisme, Théorie de la Démonstration*. Paris: Gautier-Villars, 1955.
- A. Heyting. *Intuitionism: an Introduction*. North-Holland Publishing Company, 1956.
- D. Hilbert. The logical foundations of mathematics. In *From Kant to Hilbert: A Source Book in the Foundations of Mathematics* (ed. William Ewald), volume 2. Oxford University Press (1996), 1923.
- I. Johansson. Der Minimalkalkül, ein reduzierter Intuitionistischer Formalismus. *Compositio Mathematica*, pages 119–136, 1937.
- A. Kolmogorov. On the principle of excluded middle. In *From Frege to Gödel* (ed. Jan van Heijenoort). toExcel Press (1979), 1925.
- A. Kolmogorov. On the interpretation of intuitionistic logic. In *From Brouwer To Hilbert: The Debate on the Foundations of Mathematics in the 1920s* (ed. P. Mancosu). Oxford University Press (1998), 1932.

- 
- E. Mares. Relevance logic. *The Stanford Encyclopedia of Philosophy* (Ed. E. Zalta), 2018.
- T. Nickles. From Copernicus to Ptolemy: inconsistency and method. In *Inconsistency in Science* (Ed. J. Meheus). Dordrecht: Springer, 2002.
- G. Priest and F. Berto. Dialetheism. *Stanford Encyclopedia of Philosophy*, 2013.
- N. Tennant. Natural Deduction and Sequent Calculus for Intuitionistic Relevant Logic. *The Journal of Symbolic Logic*, 52(3):665–680, 1987.
- A. S. Troelstra and D. van Dalen. *Constructivism in Mathematics: an Introduction*. Elsevier, 1988.
- M. van Atten. On the hypothetical judgement in the history of intuitionistic logic. In *Logic, Methodology, and Philosophy of Science: Proceedings of the Thirteenth International Congress* (eds. C. Glymour, W. Wei, and D. Westerstahl). College Publications, 2009.
- M. van Atten. The development of intuitionistic logic. *The Stanford Encyclopedia of Philosophy*, 2017.
- D. van Dalen. Kolmogorov and Brouwer on constructive implication and the Ex Falso rule. *Russian Mathematical Surveys*, 59:247–257, 2004.
- D. van Dalen. Another look at Brouwer’s dissertation. In *One Hundred Years of Intuitionism* (ed. M. van Atten et al.). Birkhäuser Verlag, 2008.
- D. van Dalen, editor. *The Selected Correspondence of L. E. J. Brouwer*. Springer, 2011.



# ON HOW KELSENIAN JURISPRUDENCE AND INTUITIONISTIC LOGIC HELP TO AVOID CONTRARY-TO-DUTY PARADOXES IN LEGAL ONTOLOGIES

Edward H. Haeusler<sup>1</sup> & Alexandre Rademaker<sup>2</sup>

## Abstract

In this article we show how Hans Kelsen jurisprudence and Intuitionistic logic are used to avoid the well-known contrary-to-duty (CTD) paradoxes, such as Chisholm paradoxes and its variants. This article uses an intuitionistic version of the ALC description logic, named iALC, to show how an ontology based on individually valid legal statements is able to avoid CTDs by providing models to them.

## 1. Introduction

Classical First Order Logic has been widely used as a basis for ontology creation and reasoning in many domains. These domains naturally include Legal Knowledge and Jurisprudence. As we expect, consistency is an important issue for legal ontologies. However, due to their inherently normative feature, coherence (consistency) in legal ontologies is more subtle than in other domains. Consistency, or absence of logical contradictions, seems more difficult to maintain when more than one law system can judge a case, what we call a conflict of laws. There are some legal mechanisms to solve these conflicts such as stating privileged *fori* or other ruling jurisdiction. In most of the cases, the conflict is solved by adopting a law hierarchy or precedence, rather better, ordering on laws. Even under these precedence mechanisms, coherence is still a major issue in legal systems. Each layer in this legal hierarchy has to be consistent. Since consistency is a direct consequence of how one deals with the logical negation, negation is also a main concern in legal systems. Deontic Logic, here considered as an extension of Classical Logic, has been widely used to formalize the normative aspects of the legal knowledge. There is some disagreement on using deontic logic, and any of its variants, to this task. Since a seminal paper by Alchourron and Martino (Alchourron & Martino, 1990), the propositional aspect of laws has been under discussion. In (Alchourron & Martino, 1990), the authors argue that *laws* are not to be considered as propositions, in full agreement with Hans Kelsen jurisprudence. The Kelsenian approach to Legal Ontologies considers the term “ontologies on laws” more appropriate than “law ontology”. In previous works, we showed that Classical logic is not adequate to cope with a Kelsenian based Legal Ontology. Because of the popularity of Description Logic for expressing ontologies nowadays, we developed an Intuitionistic version of Description Logic particularly devised to express Legal Ontologies. This logic is called iALC. In this article, we show how iALC avoids some Contrary-to-duty paradoxes, as *Chisholm* paradox and other paradoxes that appear in deontic logic, such as *the good samaritan* and *the knower*. For these paradoxes, we provide iALC models. Finally, we discuss the main role of the intuitionistic negation in this issue, finding out that its success may be a consequence of its paracomplete logical

---

<sup>1</sup> Pontifícia Universidade Católica do Rio de Janeiro, Departamento de Informática. E-mail: [edward.haeusler@gmail.com](mailto:edward.haeusler@gmail.com).

<sup>2</sup> Fundação Getúlio Vargas, Escola de Matemática Aplicada. E-mail: [alexandre.rademaker@fgv.br](mailto:alexandre.rademaker@fgv.br).  
DOI: 10.21452/LnA serie n.v01.book.seminario-logica-no-aviao-2013-2018.haesler-reademaker.p.44-59

aspect. This investigation opens the use of other paracomplete logics in accomplishing a logical basis for Kelsenian legal ontologies, as a complementary solution to those based on paraconsistent logics, see (M.E. Coniglio, 2009).

## 2. A brief discussion on Kelsenian Jurisprudence and its logic

A very important task in jurisprudence (legal theory) is to make precise the use of the term “law”, the *individuation problem*, and it is one of the most fundamental open questions in jurisprudence. It requires firstly answering the question “What is to count as one complete law?” (Raz, 1972). There are two main approaches to answer this question. One approach is to consider “the law” as the result of a natural process that yields a set of norms responsible for stating perfect social behavior. Another approach is to consider “the law” as a set of individual legal statements, each of them created to enforce a positively desired behavior in the society. As a consequence, in the first approach, the norms say what are the best morally speaking accepted state of affairs in a particular society, while, in the second approach, each legal statement rules an aspect of the society that the legislature wants to enforce the behavior. The first is more related to what is called *Natural Law* and the last to *Legal positivism*. We can say that the *Legal positivism* is closer to the way modeling is taken in Computer Science. In the *natural* approach to the law, it is even harder to define a system of laws than in the *legal positivism*. The natural approach demands stronger knowledge of the interdependency between the underlying legal statements than legal positivism. Because of that, the *natural* approach, in essence, is harder to be shared with practical jurisprudence principles, since they firstly are concerned to justify the law, on an essentially moral basis. This justification is quite hard to maintain from a practical point of view.

The coherence of “the law” in both approaches is essential. A debate on whether coherence is built-in by the restrictions induced by Nature in an evolutionary way, or whether coherence should be an object of knowledge management, seems to be a long debate. Despite that, legal positivism seems to be more suitable to Legal Artificial Intelligence. From the logical point of view, the *natural* approach is also harder to deal with than the *positivist* one. When describing a morally desired state-of-affairs, the logical statements take the form of propositions that has as a model best of the moral worlds. Deontic logic is suitable to be used to fulfill this task. However, a legal statement (“a law”) is essentially an individual sentence that can also be seen as an order (mandatory command), and hence, it is not a proposition at all. As a consequence, deontic logic is not appropriate to be used in knowledge bases. Besides that, (Valente, 1995) shows that deontic logic does not properly distinguish between the normative status of a situation from the normative status of a norm (rule). We think that the best jurisprudence basis for Legal ontologies and reasoning is *Legal positivism*. Thus, we will be talking a legal ontology as an ontology about (individual) laws, and not an ontology on “the law”.

Hans Kelsen initialized the *Legal positivism* tradition in 1934, for a contemporary reference see (Kelsen, 1991). He used this positive aspect of the legislature to define a theory of *pure law* and applied it to the problem of transfer citizen’s rights and obligations from one country to other when crossing borders. He produces a quite good understanding of what nowadays we denominate **Private International Law**. This achievement was so important that in many references on international law, Kelsen jurisprudence is the basis

for discussions on conflict-of-laws derived from different statements coming from different *fori*.<sup>3</sup>

In what follows we introduce the main terminology and concepts of Kelsenian jurisprudence that we use in this article. We can summarize Kelsen theory of pure law in three principles:

- (1) According to what was discussed above, individually valid legal statements are the first-class citizens of our ontology. Thus, only inhabitants of the Legal knowledge base are individual laws, see (Kelsen, 1967, supra note 5, pp 9-10)<sup>4</sup>. For example, if it is the case that Maria is married with John, and, this was legally celebrated, then “Maria-married-with-John” is an individually valid legal statement, and hence, it is a member of the Legal Ontology;
- (2) Kelsen also says that that the validity of a legal norm can only be provided concerning the validity of another, and higher, one. So,  $n_1$ , a norm, is legally valid if, and only if, it was created or promulgated in agreement with other, and higher, legally valid norm,  $n_2$ . This justification induces a precedence relationship between norms that is transitive, that is, if  $n_1$  precedes  $n_2$ , and,  $n_2$  precedes  $n_3$ , then  $n_1$  precedes  $n_3$ ;<sup>5</sup>
- (3) There is a mechanism for relating laws from one Legal system to another, the so-called “choice-of-law rule”. This mechanism is very important to the development of a concept of International Law. Assume that *Mary-is-married-with-John* is an individual legal statement in legal system *A*. Assume also that Mary is a citizen of a country adopting legal system *B*. Is there any legal statement in *B* ensuring that Mary is married in *B*? Well, this depends on *B* itself, but there is a way to connect the individual law *Mary-is-married-with-John* in *A* to *Mary-is-married-with-John* in *B*. In some legal systems, this is accomplished by what Kelsen denominated “a connection”. As shown in the following quotation from (Kelsen, 1946, page 247), the connection between the laws of *A* and *B* is made by reference, but, in fact, each law belongs to its respective legal system. In this specific case we can consider *Mary-is-married-with-John* in system *A* is connected to *Mary-is-married-with-John* in legal system *B* the connection *Lex Loci Celebrationis*.

... the law of one State prescribes the application of the law of another State, and the latter does not object or demand it. It has no right to do so since it is not really its own law which is applied by the other State. The latter applies norms of its own law. The fact that these norms have the same contents as corresponding norms of another State does not concern the latter... Since the specific technique of these norms consists in “referring” to the norms of another system and by so doing incorporating norms of identical contents into their own legal system, it would be more justifiable to call them “reference rules”... The reference

<sup>3</sup> “It is one of Kelsen’s frequently repeated doctrines that conflict of norms, in the absence of a normative procedure for resolving the conflict, shatters the concept of a unified system”, is highly emphasized in Hughes (Hughes, 1971), for example, and it is one of the principles most cited when Kelsen jurisprudence is presented.

<sup>4</sup> Kelsen takes norms and valid norms as synonyms. To say that a legal norm is valid is to say that it exists, is affirmed by Kelsen

<sup>5</sup> See (Kelsen, 1967, supra note 5, p. 196-7). This can be also found in (Kelsen, 1946, supra note 5, p. 110-1)

rule, that is ... the norm regulating the application of foreign law, may be distinguished from the norm to be applied, that is, the norm referred to. Only the former is a norm of private international law. But from a functional point of view, the one is essentially connected with the other.

Nowadays it is a common terminology in Private International Law the use of the connecting factors or legal connections between individual laws in a different legal system. Only to enumerate some of them: *Lex-Domicilii*, *Lex-Patriae*, *Lex-loci-contratum*, *Lex-loci-solutionis*, etc.

There is a philosophical problem with the principle 2 above. It demands the existence of basic laws. These basic laws do not have their validity/existence as a consequence of other more basic laws. Kelsen name these basic laws *Grundnorms*. Their validity is based on legislature acts and in a certain sense is derived from the sovereign of the State. It is out of the scope of this article to discuss such problem in Kelsen's jurisprudence. We take as granted that Kelsen jurisprudence can adequately support most of the existent legal systems, a definitively not an unreal working hypothesis.

From the three principles above, we have some very simple ontological **commitments**:

- I:** Individuals are laws;
- II:** There is a transitive and reflexive relationship between individual laws that reflects the natural precedence relationship between laws;
- III:** There are legal connections between individual laws in different legal systems or between different *fori* in the same broader legal system.

From these commitments, we derive the basic constructs of the logic iALC. In the first place, our legal ontology relates concepts to legal systems. Description logics uses nominals to refer to individuals. So, an expression as  $i : A$ , stands for  $i$  is an individual law, belonging to the legal system  $A$ , a concept.

From commitment we consider an expression as  $i \preceq j$  standing for the individual law  $i$  legally precedes individual law  $j$ . The subsumption relationship  $A \subseteq B$ , from description logic, denotes that  $A$  is a legal subsystem of  $B$ . One could interpret this relation as the inclusion relationship.<sup>6</sup> We discuss the implications of using negated contents together with Kelsenian jurisprudence in the following. This can be found in (Haeusler, de Paiva, & Rademaker, 2011, 2010; Haeusler, Paiva, & Rademaker, 2010; Haeusler, de Paiva, & Rademaker, 2010) too.

Under the classical setting, a negated concept  $\neg A$  denotes the **set** of all inhabitants of the domain that do not belong to the interpretation of  $A$ . Under ontological commitment there is no individual law that does not exist in, belong to, the domain. Since norms and laws are not propositions, it is a complete nonsense to negate a law. As we already seen, we can negate a concept on laws. Consider the collection of all Brazilian individual laws. Call it  $BR$ . In a classical setting  $BR \sqcup \neg BR$  is the universe of laws. Thus, any law that it is not in  $BR$  has to be a law outside  $BR$ , that is, belonging to  $\neg BR$ . For example, if Peter is 17 years old, it is not liable according to the Brazilian law. Is PETERISLIABLE a valid law at all? If so, it has to belong to  $\neg BR$ . Using Kelsen in a classical setting, individual laws not belonging to a concept automatically belong to its complementary concept. The problem with this is that it is possible to create laws outside a jurisdiction or forum by the very simple act of considering or experimenting a legal situation. Nowadays in Brazil, the parliament is discussing the liability under the 16 years. By the simple fact of discussing

<sup>6</sup> In Classical ALC this is just the case, but we shown here that classical reasoning it is not a good choice for dealing with legal ontologies

the validity of their corresponding individual laws, we are forced to accept they exist outside the Brazilian legal system. We do not consider this feature appropriate to legal ontology definition. Dealing with negations every time we assume the existence of a law may bring unnecessary complexity to legal ontology definition. Because the precedence relationship between laws, cf. ontological commitment , there is a natural alternative to classical logic, the intuitionistic logic (IL). According to IL semantics,  $i : \neg A$ , iff, for each law  $j$ , such that  $i \preceq j$ , it is not the case that  $j : A$ . This semantics means that  $i$  does not provide any legal support for any individual law belong to  $A$ , which agrees with Kelsen jurisprudence on the hierarchy of individual laws.

Commitment gives rise to expressions of the form  $m$  LEXLOCICELEBRATIONIS  $m$ , where  $m$  is MARYISMARRIEDWITHJOHN and LEXLOCICELEBRATIONIS is a legal connection. Thus, if *Abroad* is the concept that represents all laws in Portugal, then the concept  $\exists$  LEXLOCICELEBRATIONIS *Portugal* represents the Brazilian individual laws stating that Portuguese marriage is valid in Brazil. The private international law of any country is a collection of laws stated in similar ways for every possible legal connection. In (Haeusler, de Paiva, & Rademaker, 2010) it is shown in detail a judicial case deriving that a renting contract is solving a conflict of laws in space through private international law.

### 3. Some philosophical discussion on using Kelsen in legal ontologies

We base our work on two ontological criteria: <sup>7</sup>

- Ontological Commitment (due to W.Quine), our logical approach is ontologically committed to Valid Legal Statements only, as discussed in section . The only nominals occurring in our logic language are valid individual laws, and;
- Ontological Parsimony, which is strongly related to Quine’s ontological commitment too, with a mention of its stronger version also known as Occam’s Razor, here denoted as OR.

The second criteria is based on: “One ‘easy’ case where OR can be straightforwardly applied is when a theory  $T$ , postulates entities which are explanatorily idle. Excising these entities from  $T$  produces a second theory,  $T^*$ , which has the same theoretical virtues as  $T$  but a smaller set of ontological commitments. Hence, according to OR, it is rational to pick  $T^*$  over  $T$ .”

We observe that nominals, representing individuals, denote only valid individual laws and nothing in the iALC language described in the following section, is committed with non-valid individual laws, according to the second ontological criterion above, we do not have to consider non-valid individual laws. Technically speaking there is no element in the iALC language able to denote an invalid individual law in any model of any iALC theory. If something is a valid individual law regarded some legal system in some place in the world, then this individual belongs to our semantic universe.

This philosophical basis allows us to have only sets of valid individuals as semantics for iALC theories. Thus, as the a reviewer have already observed, this implies that  $\neg A$  is the set of individual laws holding outside Brazil, and the classical negation is not adequate to denote this set. If we get  $\neg A$  meaning “individual laws that *do not* hold in Brazil”, the set of laws being a proper subset of the universe, and  $A$  is the conjunctive property “laws + holds in  $A$ ”. Then the complement,  $\neg A$  would be all elements of the universe which are

<sup>7</sup> see Quine’s “On What there is” article and <http://plato.stanford.edu/entries/simplicity>, for example to a primer ontological criteria

either not a valid individual law or do not hold in Brazil. But there is no way to take the semantics in this way, for the semantics we get from the ontological commitment from section is given by “The individual valid laws holding outside of Brazil”.

Finally, concerning contradictory individual laws, they can coexist in the same universe, since they are there because they hold in distinct legal systems. In fact they are apparently contradictory. For example, “There is death penalty” and “Death sentence is not allowed” can coexist, since there are countries where each of these legal statements are valid. Concretely: “There is death penalty”:Iran and “Death sentence is not allowed”:Brazil.

#### 4. The Logic iALC

Classical Description Logic has been widely used as a basis for ontology creation and reasoning in many knowledge specific domains, including Legal AI.

An adequate intuitionistic semantics for negation in a legal domain comes to the fore when we take legally valid individual statements as the inhabitants of our legal ontology. This allows us to elegantly deal with particular situations of legal coherence, such as conflict of laws, as those solved by Private International Law analysis. In (Haeusler, de Paiva, & Rademaker, 2010, 2010; Haeusler, Paiva, & Rademaker, 2010) we present an Intuitionistic Description Logic, called iALC for Intuitionistic ALC (for Attributive Language with Complements, the canonical classical description logic system). A labeled sequent calculus for iALC based on a labeled sequent calculus for ALC (Rademaker, 2012), was also presented. In these previous articles, we discussed the jurisprudence foundation of our system, and show how we can perform a coherence analysis of “Conflict of Laws in Space” by means of iALC. This conflict happens when several laws can be applied, with different outcomes, to a case depending on the place where the case occurs. Typical examples are those ruling the rights of a citizen abroad.

In (Haeusler, de Paiva, & Rademaker, 2010), the semantics of iALC is precisely provided and follows the framework for constructive modal logics presented by Simpson (Simpson, 1993) and adapted to description languages by Paiva (de Paiva, 2006). In the cited reference, we applied iALC to the problem of formalizing legal knowledge.

Description Logics are an important knowledge representation formalism, unifying and giving a logical basis to the well known AI frame-based systems of the eighties. Description logics are very popular right now. Given the existent and proposed applications of the Semantic Web, there has been a fair amount of work into finding the most well-behaved system of description logic that has the broadest application, for any specific domain. Description logics tend to come in families of logical systems, depending on which concept constructors you allow in the logic. Since description logics came into existence as fragments of first-order logic chosen to find the best trade-off possible between expressiveness and tractability of the fragment, several systems were discussed and in the taxonomy of systems that emerged the ALC has come to be known as the canonical one. The basic building blocks of description logics are *concepts*, *roles* and *individuals*. Concepts are described as unary predicates in usual first-order logic and roles as binary atomic predicates used to modify the concepts.

As discussed in (de Paiva, 2006), considering versions of *constructive* description logics makes sense, both from a theoretical and from a practical viewpoint. There are several possible and sensible ways of defining *constructive* description logics, whether your motivation is natural language semantics, (de Paiva, 2006), or Legal AI, (Haeusler, de Paiva,

& Rademaker, 2010). As far as *constructive* description logics are concerned, Mendler and Scheele have worked out a very compelling system *cALC* (Mendler & Scheele, 2010), based on the constructive modal logic *CK* (Bellin, de Paiva, & Ritter, 2001)), one possible choice for us. However in this note we follow a different path and describe a constructive version of *ALC*, based on the framework for constructive modal logics developed by Simpson (the system *IK*) in his phd thesis (Simpson, 1993) (For a proof-theoretic comparison between the constructive modal logics *CK* and *IK* one can see (Ranalter, 2010)).

Our motivation, besides Simpson’s work, is the framework developed by Braüner and de Paiva in (Braüner & de Paiva, 2006) for constructive Hybrid Logics. We reason that having already frameworks for constructive modal and constructive hybrid logics in the labelled style of Simpson, we might end up with the best style of constructive description logics, in terms of both solid foundations and ease of implementation. Since submitting this paper we have been told about the master thesis of Clément (Clément, 2008) which follows broadly similar lines. Clément proves soundness and completeness of this system and then provides a focused version of it, a very interesting development, as focused systems are, apparently, very useful for proof search.

Building up from the Simpson’s constructive modal logics in (Braüner & de Paiva, 2006) (called here *IML*), it is introduced *intuitionistic hybrid logics*, denoted by *IHL*. Hybrid logics add to usual modal logics a new kind of propositional symbols, the *nominals*, and also the so-called *satisfaction operators*. A nominal is assumed to be true at exactly one world, so a nominal can be considered the name of a world. If  $x$  is a nominal and  $X$  is an arbitrary formula, then a new formula  $x:X$  called a satisfaction statement can be formed. The satisfaction statement  $x:X$  expresses that the formula  $X$  is true at one particular world, namely the world denoted by  $x$ . In hindsight one can see that *IML* shares with hybrid formalisms the idea of making the possible-world semantics part of the deductive system. While *IML* makes the relationship between worlds (e.g.,  $xRy$ ) part of the deductive system, *IHL* goes one step further and sees the worlds themselves  $x, y$  as part of the deductive system, (as they are now nominals) and the satisfaction relation itself as part of the deductive system, as it is now a syntactic operator, with modality-like properties.

Our Sequent Calculus for *iALC* was first presented in (de Paiva, Hausler, & Rademaker, 2010) where we briefly described the immediate properties of this system and most importantly we discuss a case study of the use of *iALC* in legal AI.

A very important observation is that this article corrects and extends the presentation of *iALC* appearing in all previous articles. It points out the difference between *iALC* and the intuitionistic hybrid logic presented in (de Paiva, 2006). Completeness and soundness proofs are revised. A discussion on the computational complexity of *iALC* is also taken.

## 5. Intuitionistic ALC

The *iALC* logic is based on the framework for intuitionistic modal logic *IK* proposed in (Simpson, 1993; Fischer-Servi, 1984; Plotkin & Stirling, 1986). These modal logics arise from interpreting the usual possible worlds definitions in an intuitionistic meta-theory. As we will see in the following paragraphs, ideas from (Braüner & de Paiva, 2006) were also used, where the framework *IHL*, for *intuitionistic hybrid logics*, is introduced. *iALC* concepts are described as:

$$C, D ::= A \mid \perp \mid \top \mid \neg C \mid C \sqcap D \mid C \sqcup D \mid C \sqsubseteq D \mid \exists R.C \mid \forall R.C$$

where  $C, D$  stands for concepts,  $A$  for an atomic concept,  $R$  for an atomic role. We could have used distinct symbols for subsumption of concepts and the subsumption concept constructor but this would blow-up the calculus presentation. This syntax is more general than standard ALC since it includes subsumption  $\sqsubseteq$  as a concept-forming operator. We have no use for nested subsumptions, but they do make the system easier to define, so we keep the general rules. Negation could be defined via subsumption, that is,  $\neg C = C \sqsubseteq \perp$ , but we find it convenient to keep it in the language. The constant  $\top$  could also be omitted since it can be represented as  $\neg\perp$ .

A constructive interpretation of iALC is a structure  $\mathcal{I}$  consisting of a non-empty set  $\Delta^{\mathcal{I}}$  of entities in which each entity represents a partially defined individual; a refinement pre-ordering  $\preceq^{\mathcal{I}}$  on  $\Delta^{\mathcal{I}}$ , i.e., a reflexive and transitive relation; and an interpretation function  $\cdot^{\mathcal{I}}$  mapping each role name  $R$  to a binary relation  $R^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}} \times \Delta^{\mathcal{I}}$  and atomic concept  $A$  to a set  $A^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}}$  which is closed under refinement, i.e.,  $x \in A^{\mathcal{I}}$  and  $x \preceq^{\mathcal{I}} y$  implies  $y \in A^{\mathcal{I}}$ . The interpretation  $\mathcal{I}$  is lifted from atomic concepts to arbitrary concepts via:

$$\begin{aligned} \top^{\mathcal{I}} &=_{df} \Delta^{\mathcal{I}} \\ \perp^{\mathcal{I}} &=_{df} \emptyset \\ (\neg C)^{\mathcal{I}} &=_{df} \{x \mid \forall y \in \Delta^{\mathcal{I}}. x \preceq y \Rightarrow y \notin C^{\mathcal{I}}\} \\ (C \sqcap D)^{\mathcal{I}} &=_{df} C^{\mathcal{I}} \cap D^{\mathcal{I}} \\ (C \sqcup D)^{\mathcal{I}} &=_{df} C^{\mathcal{I}} \cup D^{\mathcal{I}} \\ (C \sqsubseteq D)^{\mathcal{I}} &=_{df} \{x \mid \forall y \in \Delta^{\mathcal{I}}. (x \preceq y \text{ and } y \in C^{\mathcal{I}}) \Rightarrow y \in D^{\mathcal{I}}\} \\ (\exists R.C)^{\mathcal{I}} &=_{df} \{x \mid \exists y \in \Delta^{\mathcal{I}}. (x, y) \in R^{\mathcal{I}} \text{ and } y \in C^{\mathcal{I}}\} \\ (\forall R.C)^{\mathcal{I}} &=_{df} \{x \mid \forall y \in \Delta^{\mathcal{I}}. x \preceq y \Rightarrow \forall z \in \Delta^{\mathcal{I}}. (y, z) \in R^{\mathcal{I}} \Rightarrow z \in C^{\mathcal{I}}\} \end{aligned}$$

Following the semantics of IK, the structures  $\mathcal{I}$  are models for iALC if they satisfy two frame conditions:

**F1:** if  $w \leq w'$  and  $wRv$  then  $\exists v'. w'Rv'$  and  $v \leq v'$

**F2:** if  $v \leq v'$  and  $wRv$  then  $\exists w'. w'Rv'$  and  $w \leq w'$

The above conditions are diagrammatically expressed as:

$$\begin{array}{ccc} w' & \xrightarrow{R} & v' \\ \uparrow & (F1) & \uparrow \\ w & \xrightarrow{R} & v \end{array} \quad \text{and} \quad \begin{array}{ccc} w' & \xrightarrow{R} & v' \\ \uparrow & (F2) & \uparrow \\ w & \xrightarrow{R} & v \end{array}$$

Our setting simplifies (Mendler & Scheele, 2010), since *iALC* satisfies (like classical ALC)  $\exists R.\perp = \perp$  and  $\exists R.(C \sqcup D) = \exists R.C \sqcup \exists R.D$ .

In contrast with the above mentioned approaches, ours assign a truth values to some formulas, also called assertions, they are not concepts as in (Bräuner & de Paiva, 2006), for example. Below we define the syntax of general assertions ( $A$ ) and nominal assertions ( $N$ ) for ABOX reasoning in iALC. Formulas ( $F$ ) also includes subsumption of concepts interpreted as propositional statements.

$$N ::= x : C \mid x : N \qquad A ::= N \mid xRy \qquad F ::= A \mid C \sqsubseteq C$$

where  $x$  and  $y$  are nominals,  $R$  is a role symbol and  $C$  is a concept. In particular, this allows  $x : (y : C)$ , which is a perfectly valid nominal assertion.

**DEFINITION 1.** (outer nominal) *In a nominal assertion  $x : \gamma$ ,  $x$  is said to be the outer nominal of this assertion. That is, in an assertion of the form  $x : (y : \gamma)$ ,  $x$  is the outer nominal.*



We write  $\mathcal{I}, w \models C$  to abbreviate  $w \in C^{\mathcal{I}}$  which means that entity  $w$  satisfies concept  $C$  in the interpretation  $\mathcal{I}$ <sup>8</sup>. Further,  $\mathcal{I}$  is a model of  $C$ , written  $\mathcal{I} \models C$  iff  $\forall w \in \mathcal{I}$ , we have that  $\mathcal{I}, w \models C$ . Finally,  $\models C$  means  $\forall \mathcal{I}$ , we have that  $\mathcal{I} \models C$ . All previous notions are extended to sets  $\Phi$  of concepts in the usual universal fashion. Given the hybrid satisfaction statements, the interpretation and semantic satisfaction relation are extended in the expected way. The statement  $\mathcal{I}, w \models x : C$  holds, if and only if,  $\forall z_x \succeq^{\mathcal{I}} x$ , we have that,  $\mathcal{I}, z_x \models C$ . In a similar fashion,  $\mathcal{I}, w \models xRy$  holds, if and only if,  $\forall z_x \succeq x. \forall z_y \succeq y. (x_x^{\mathcal{I}}, z_y^{\mathcal{I}}) \in R^{\mathcal{I}}$ . That is, the evaluation of the hybrid formulas does not take into account only the world  $w$ , but it has to be monotonically preserved. It can be observed that for every  $w'$ , if  $x^{\mathcal{I}} \preceq w'$  and  $\mathcal{I}, x' \models \alpha$ , then  $\mathcal{I}, w' \models \alpha$  is a property holding on this satisfaction relation.

In common reasoning tasks the interpretation  $\mathcal{I}$  and the entity  $w$  in a verification goal such as  $\mathcal{I}, w \models \delta$  are not given directly but are themselves axiomatized by sets of concepts and formulas. Usually we have a set  $\Theta$ <sup>9</sup> of formulas and the set  $\Gamma$  of concepts. Accordingly:

**DEFINITION 2.** *We write  $\Theta, \Gamma \models \delta$  if it is the case that:*

$$\forall \mathcal{I}. \text{ if } (\forall x \in \Delta^{\mathcal{I}}. (\mathcal{I}, x \models \Theta)), \text{ then } \forall (Nom(\Gamma, \delta)). \forall \vec{z} \succeq Nom(\Gamma, \delta). (\mathcal{I}, \vec{z} \models \Gamma \Rightarrow \mathcal{I}, \vec{z} \models \delta) \quad (1)$$

where  $\vec{z}$  denotes a vector of variables  $z_1, \dots, z_k$  and  $Nom(\Gamma, \delta)$  is the vector of all outer nominals occurring in each nominal assertion of  $\Gamma \cup \{\delta\}$ .  $x$  is the only outer nominal of a nominal assertion  $\{x : \gamma\}$ , while a (pure) concept  $\gamma$  has no outer nominal.

A Hilbert calculus for iALC is provided following (Plotkin & Stirling, 1986; Simpson, 1993; Fischer-Servi, 1984). It consists of all axioms of intuitionistic propositional logic plus the axioms and rules displayed in Figure 1. The Hilbert calculus implements TBox-reasoning. That is, it decides the semantical relationship  $\Theta, \emptyset \models C$ .  $\Theta$  has only formulas as members.

0. all substitution instances of theorems of IPL
  1.  $\forall R.(C \sqsubseteq D) \sqsubseteq (\forall R.C \sqsubseteq \forall R.D)$
  2.  $\exists R.(C \sqsubseteq D) \sqsubseteq (\exists R.C \sqsubseteq \exists R.D)$
  3.  $\exists R.(C \sqcup D) \sqsubseteq (\exists R.C \sqcup \exists R.D)$
  4.  $\exists R.\perp \sqsubseteq \perp$
  5.  $(\exists R.C \sqsubseteq \forall R.C) \sqsubseteq \forall R.(C \sqsubseteq D)$
- MP** If  $C$  and  $C \sqsubseteq D$  are theorems,  $D$  is a theorem too.
- Nec** If  $C$  is a theorem then  $\forall R.C$  is a theorem too.

**FIGURE 1.** The iALC axiomatization

A Sequent Calculus for iALC is also provided. The logical rules of the Sequent Calculus for iALC are presented in Figure 2.<sup>10</sup> The structural rules and the cut rule are omitted but they are as usual. The  $\delta$  stands for concepts or assertions ( $x : C$  or  $xRy$ ),  $\alpha$  and  $\beta$

<sup>8</sup> In IHL, this  $w$  is a world and this satisfaction relation is possible world semantics

<sup>9</sup> Here we consider only acycled TBox with  $\sqsubseteq$  and  $\equiv$ .

<sup>10</sup> The reader may want to read Proof Theory books, for example, (Takeuti, 2013; Buss, 1998; Negri & Von Plato, 2008; Girard, Taylor, & Lafont, 1989).

for concept and  $R$  for role.  $\Delta$  is a set of formulas. In rules p- $\exists$  and p- $\forall$ , the syntax  $\forall R.\Delta$  means  $\{\forall R.\alpha \mid \alpha \in \text{concepts}(\Delta)\}$ , that is, all concepts in  $\Delta$  are universal quantified with the same role. The assertions in  $\Delta$  are kept unmodified. In the same way, in rule p-N the addition of the nominal is made only in the concepts of  $\Delta$  (and in  $\delta$  if that is a concept) keeping the assertions unmodified.

The propositional connectives ( $\sqcap, \sqcup, \sqsubseteq$ ) rules are as usual, the rule  $\sqcup_2$ -r is omitted. The rules are presented without nominals but for each of these rules there is a counterpart with nominals. For example, the rule  $\sqsubseteq$ -r has one similar:

$$\frac{\Delta, x: \alpha \Rightarrow x: \beta}{\Delta \Rightarrow x: (\alpha \sqsubseteq \beta)} \text{n-}\sqsubseteq\text{-r}$$

The main modification comes for the modal rules, which are now role quantification rules. We must keep the intuitionistic constraints for modal operators. Rule  $\exists$ -l has the usual condition that  $y$  is not in the conclusion. Concerning the usual condition on the  $\forall$ -r rule, it is not the case in this system, for the interpretation of the a nominal assertion in a sequent is already implicitly universal (Definition 2).

$$\begin{array}{c} \overline{\Delta, \delta \Rightarrow \delta} \\ \frac{\Delta, xRy \Rightarrow y: \alpha}{\Delta \Rightarrow x: \forall R.\alpha} \forall\text{-r} \\ \frac{\Delta \Rightarrow xRy \quad \Delta \Rightarrow y: \alpha}{\Delta \Rightarrow x: \exists R.\alpha} \exists\text{-r} \\ \frac{\Delta, \alpha \Rightarrow \beta}{\Delta \Rightarrow \alpha \sqsubseteq \beta} \sqsubseteq\text{-r} \\ \frac{\Delta \Rightarrow \alpha \quad \Delta \Rightarrow \beta}{\Delta \Rightarrow \alpha \sqcap \beta} \sqcap\text{-r} \\ \frac{\Delta \Rightarrow \alpha}{\Delta \Rightarrow \alpha \sqcup \beta} \sqcup_1\text{-r} \\ \frac{\Delta, \alpha \Rightarrow \beta}{\forall R.\Delta, \exists R.\alpha \Rightarrow \exists R.\beta} \text{p-}\exists \\ \frac{\Delta \Rightarrow \delta}{x: \Delta \Rightarrow x: \delta} \text{p-N} \\ \overline{\Delta, x: \perp \Rightarrow \delta} \\ \frac{\Delta, x: \forall R.\alpha, y: \alpha, xRy \Rightarrow \delta}{\Delta, x: \forall R.\alpha, xRy \Rightarrow \delta} \forall\text{-l} \\ \frac{\Delta, xRy, y: \alpha \Rightarrow \delta}{\Delta, x: \exists R.\alpha \Rightarrow \delta} \exists\text{-l} \\ \frac{\Delta_1 \Rightarrow \alpha \quad \Delta_2, \beta \Rightarrow \delta}{\Delta_1, \Delta_2, \alpha \sqsubseteq \beta \Rightarrow \delta} \sqsubseteq\text{-l} \\ \frac{\Delta, \alpha, \beta \Rightarrow \delta}{\Delta, \alpha \sqcap \beta \Rightarrow \delta} \sqcap\text{-l} \\ \frac{\Delta, \alpha \Rightarrow \delta \quad \Delta, \beta \Rightarrow \delta}{\Delta, \alpha \sqcup \beta \Rightarrow \delta} \sqcup\text{-l} \\ \frac{\Delta \Rightarrow \alpha}{\forall R.\Delta \Rightarrow \forall R.\alpha} \text{p-}\forall \end{array}$$

**FIGURE 2.** The System  $\text{SC}_{\text{iALC}}$ : logical rules

**THEOREM 1.** *The sequent calculus described in Fig. 2 is sound and complete for TBox reasoning, that is  $\Theta, \emptyset \models C$  if and only if  $\Theta \Rightarrow C$  is derivable with the rules of Figure 2.*

The completeness of our system is proved relative to the axiomatization of  $\text{iALC}$ , shown in Figure 1. The proof is presented in Section .

The soundness of the system is proved directly from the semantics of  $\text{iALC}$  including the ABOX, that is, including nominals. The semantics of a sequent is defined by the satisfaction relation, as shown in Definition 2. The sequent  $\Theta, \Gamma \Rightarrow \delta$  is valid if and only

if  $\Theta, \Gamma \models \gamma$ . Soundness is proved by showing that each sequent rule preserves the validity of the sequent and that the initial sequent is valid. This proof is presented in Section .

We note that although we have here fixed some inaccuracies in the presentation of the  $\text{iALC}$  semantics in (de Paiva et al., 2010), the system presented here is basically the same, excepted that here the propositional rules are presented without nominals. Given that, the soundness of the system proved in (de Paiva et al., 2010) can be still considered valid without further problems. Note also that the proof of soundness provides in Section is regarded the full language of  $\text{iALC}$ . It considers nominals and assertion on nominals relationship, that is it concerns ABOX and TBOX. The proof of completeness is for the TBOX only. A proof of completeness for ABOX can be done by the method of canonical models. For the purposes of this article, we choose to show the relative completeness proof with the sake of showing a simpler proof concerning TBOX.

### 6. The completeness of $\text{SC}_{\text{iALC}}$ system

We show the relative completeness of  $\text{SC}_{\text{iALC}}$  regarding the axiomatic presentation of  $\text{iALC}$  presented in Figure 1. To prove the completeness of  $\text{SC}_{\text{iALC}}$  it is sufficient to derive in  $\text{SC}_{\text{iALC}}$  the axioms 1–5 of  $\text{iALC}$ . It is clear that all substitution instances of IPL theorems can also be proved in  $\text{SC}_{\text{iALC}}$  using only propositional rules. The MP rule is a derived rule from the  $\text{SC}_{\text{iALC}}$  using the cut rule. The Nec rule is the  $\text{p-}\forall$  rule in the system with  $\Delta$  empty. In the first two proofs below do not use nominals for given better intuition of the reader about the use of rules with and without nominals.

Axiom 1:

$$\frac{\frac{\frac{\alpha \Rightarrow \alpha \quad \beta \Rightarrow \beta}{\alpha \sqsubseteq \beta, \alpha \Rightarrow \beta} \sqsubseteq\text{-l}}{\forall R.(\alpha \sqsubseteq \beta), \exists R.\alpha \Rightarrow \exists R.\beta} \text{p-}\exists}{\forall R.(\alpha \sqsubseteq \beta) \Rightarrow \exists R.\alpha \sqsubseteq \exists R.\beta} \sqsubseteq\text{-r}$$

Axiom 2:

$$\frac{\frac{\frac{\alpha \Rightarrow \alpha \quad \beta \Rightarrow \beta}{\alpha \sqsubseteq \beta, \alpha \Rightarrow \beta} \sqsubseteq\text{-l}}{\forall R.(\alpha \sqsubseteq \beta), \forall R.\alpha \Rightarrow \forall R.\beta} \text{p-}\forall}{\forall R.(\alpha \sqsubseteq \beta) \Rightarrow \forall R.\alpha \sqsubseteq \forall R.\beta} \sqsubseteq\text{-r}$$

Axiom 3:

$$\frac{\frac{xRy, y : \perp \Rightarrow x : \perp}{x : \exists R.\perp \Rightarrow x : \perp} \exists\text{-l}}{\Rightarrow x : (\exists R.\perp \sqsubseteq \perp)} \sqsubseteq\text{-r}$$

Axiom 4:

$$\frac{\frac{x : \exists R.\alpha \Rightarrow x : \exists R.\alpha}{x : \exists R.\alpha \Rightarrow x : (\exists R.\alpha \sqcup \exists R.\beta)} \sqcup\text{-l-r} \quad \frac{x : \exists R.\beta \Rightarrow x : \exists R.\beta}{x : \exists R.\beta \Rightarrow x : (\exists R.\alpha \sqcup \exists R.\beta)} \sqcup\text{-r}}{x : \exists R.(\alpha \sqcup \beta) \Rightarrow x : (\exists R.\alpha \sqcup \exists R.\beta)} \sqcup\text{-l}$$

Axiom 5:

$$\frac{\frac{xRy, y : \alpha \Rightarrow y : \alpha \quad xRy, y : \alpha \Rightarrow xRy}{xRy, y : \alpha \Rightarrow x : \exists R.\alpha} \exists\text{-r} \quad \frac{xRy, y : \alpha, y : \beta, \forall R.\beta \Rightarrow y : \beta}{xRy, y : \alpha, x : \forall R.\beta \Rightarrow y : \beta} \forall\text{-l}}{\frac{\frac{x : (\exists R.\alpha \sqsubseteq \forall R.\beta), xRy, y : \alpha \Rightarrow y : \beta}{x : (\exists R.\alpha \sqsubseteq \forall R.\beta), xRy \Rightarrow y : (\alpha \sqsubseteq \beta)} \forall\text{-r}}{x : (\exists R.\alpha \sqsubseteq \forall R.\beta) \Rightarrow x : \forall R.(\alpha \sqsubseteq \beta)} \forall\text{-r}}{\Rightarrow x : [(\exists R.\alpha \sqsubseteq \forall R.\beta) \sqsubseteq \forall R.(\alpha \sqsubseteq \beta)]} \sqsubseteq\text{-r}$$

## 7. Soundness of $\text{SC}_{\text{iALC}}$ system

In this section we prove that.

**PROPOSITION 1.** *If  $\Theta, \Gamma \Rightarrow \delta$  is provable in  $\text{SC}_{\text{iALC}}$  then  $\Theta, \Gamma \models \gamma$ .*

*Proof:* We prove that each sequent rule preserves the validity of the sequent and that the initial sequents are valid. The definition of a valid sequent ( $\Theta, \Gamma \models \gamma$ ) is presented in Definition 2.

The validity of the axioms is trivial. We first observe that any application of the rules  $\sqsubseteq\text{-r}$ ,  $\sqsubseteq\text{-l}$ ,  $\sqcap\text{-r}$ ,  $\sqcap\text{-l}$ ,  $\sqcup\text{-r}$ ,  $\sqcup\text{-l}$  of  $\text{SC}_{\text{iALC}}$  where the sequents do not have any nominal, neither in  $\Theta$  nor in  $\Gamma$ , is sound regarded intuitionistic propositional logic Kripke semantics, to which the validity definition above collapses whenever there is no nominal in the sequents. Thus, in this proof we concentrate in the case where there are nominals. We first observe that the nominal version of  $\sqsubseteq\text{-r}$ , the validity of the premises includes

$$\forall(\text{Nom}(\Gamma, \delta)). \forall \vec{z} \succeq \text{Nom}(\Gamma, \delta). (\mathcal{I}, \vec{z} \models \Gamma \Rightarrow \mathcal{I}, \vec{z} \models \delta)$$

This means that  $\Gamma$  holds in any worlds  $\vec{z} \succeq \vec{x}$  for the vector  $\vec{x}$  of nominals occurring in  $\Gamma$ . This includes the outer nominal  $x_i$  in  $\delta$  (if any). In this case the semantics of  $\sqsubseteq$  is preserved, since  $\vec{z}$  includes  $z_i \succeq x_i$ . With the sake of a more detailed analysis, we consider the following instance:

$$\frac{x : \alpha_1, y : \alpha_2 \Rightarrow x : \beta}{\alpha_1 \Rightarrow x : \alpha_2 \sqsubseteq \beta} \sqsubseteq\text{-r}$$

Consider an  $\text{iALC}$  structure  $\mathcal{I} = \langle \mathcal{U}, \preceq, R^{\mathcal{I}}, \dots, C^{\mathcal{I}} \rangle$ . In this case, for any  $\mathcal{I}$  and any  $z_1, z_2 \in \mathcal{U}^{\mathcal{I}}$  if  $z_1 \succeq x^{\mathcal{I}}$ ,  $z_1 \succeq y^{\mathcal{I}}$ , such that,  $\mathcal{I}, z_i \models \alpha_1$  and  $\mathcal{I}, z_i \models \alpha_2$ , we have that  $\mathcal{I}, z_i \models x : \beta$ , since the premise is valid, by hypothesis. In this case, by the semantics of  $\sqsubseteq$  we have  $\mathcal{I}, z_i \models x : \alpha_1 \sqsubseteq \beta$ . The conclusion of the rule is valid too.

The argument shown above for the  $\sqsubseteq\text{-r}$  rule is analogous for the nominal versions of  $\sqsubseteq\text{-r}$ ,  $\sqsubseteq\text{-l}$ ,  $\sqcap\text{-r}$ ,  $\sqcap\text{-l}$ ,  $\sqcup\text{-r}$ ,  $\sqcup\text{-l}$ . Consider the rule  $\forall\text{-r}$ .

$$\frac{\Delta, xRy \Rightarrow y : \alpha}{\Delta \Rightarrow x : \forall R.\alpha} \forall\text{-r}$$

Since the premise is valid we have that if  $\forall z_x \succeq x^{\mathcal{I}}, \forall z_y \succeq y^{\mathcal{I}}, (z_x, z_y) \in R^{\mathcal{I}}$  then  $\forall z_y \succeq y^{\mathcal{I}}, \mathcal{I}, z_y \models \gamma$ . This entails that  $x^{\mathcal{I}} \in (\forall R.\gamma)^{\mathcal{I}}$ , for  $x^{\mathcal{I}} \succeq x^{\mathcal{I}}$ . We observe that by the restriction on the rule application,  $y$  does not occur in  $\Delta$ , it only occurs in  $xRy$  and  $y : \alpha$ . The truth of these formulas are subsumed by  $\forall R.\gamma$ . The conclusion does not need to consider them any more. The conclusion is valid too. Another way to see its soundness is to prove that if  $xRy \Rightarrow y : \alpha$  is valid, then so is  $\Rightarrow x : \forall R.\alpha$ . This can be show by the following reasoning:

$$\forall x^{\mathcal{I}} \forall y^{\mathcal{I}} \forall z_x \forall z_y (z_x \succeq x^{\mathcal{I}} \rightarrow (z_y \succeq y^{\mathcal{I}} \rightarrow ((z_x, z_y) \in R^{\mathcal{I}} \rightarrow \mathcal{I}, z_y \models y : \alpha)))$$

that is the same as:

$$\forall x^{\mathcal{I}} \forall y^{\mathcal{I}} \forall z_x \forall z_y (z_x \succeq x^{\mathcal{I}} \rightarrow (z_y \succeq y^{\mathcal{I}} \rightarrow ((z_x, z_y) \in R^{\mathcal{I}} \rightarrow \mathcal{I}, y^{\mathcal{I}} \models \alpha)))$$

Using the fact that  $\forall y^{\mathcal{I}} (y^{\mathcal{I}} \succeq y^{\mathcal{I}})$ , we obtain:

$$\forall x^{\mathcal{I}} \forall z_x (z_x \succeq x^{\mathcal{I}} \rightarrow \forall y^{\mathcal{I}} ((z_x, y^{\mathcal{I}}) \in R^{\mathcal{I}} \rightarrow \mathcal{I}, y^{\mathcal{I}} \models \alpha))$$

The above condition states that  $\Rightarrow x : \forall R.\alpha$  is valid.

$$\forall x^{\mathcal{I}} \forall y^{\mathcal{I}} \forall z_x \forall z_y (z_x \succeq x^{\mathcal{I}} \rightarrow (z_y \succeq y^{\mathcal{I}} \rightarrow ((z_x, z_y) \in R^{\mathcal{I}} \rightarrow \mathcal{I}, z_y \models y : \alpha)))$$

Consider the rule  $\forall\text{-l}$ :

$$\frac{\Delta, x : \forall R.\alpha, y : \alpha, xRy \Rightarrow \delta}{\Delta, x : \forall R.\alpha, xRy \Rightarrow \delta} \forall\text{-l}$$

As in the  $\forall\text{-r}$  case, we analyze the simplest validity preservation: if  $x : \forall R.\alpha \wedge xRy$  is valid, then so is  $x : \forall R.\alpha \wedge y : \alpha \wedge xRy$ . The first condition is:

$$\begin{aligned} \forall x^{\mathcal{I}} \forall y^{\mathcal{I}} \forall z_x (z_x \succeq x^{\mathcal{I}} \rightarrow \forall z_y (z_y \succeq y^{\mathcal{I}} \rightarrow \\ ((\mathcal{I}, z_y \models x : \forall R.\alpha) \wedge (\mathcal{I}, z_y \models x : \forall R.\alpha) \wedge ((z_x, z_y) \in R^{\mathcal{I}}) \rightarrow \\ (\mathcal{I}, z_y \models y : \alpha) \wedge (\mathcal{I}, z_x \models y : \alpha)))) \end{aligned} \quad (2)$$

Using  $z_y = y^{\mathcal{I}}$ , eliminating  $z_x$  from the term, and, using the fact that  $\mathcal{I}, z_y \models y : \alpha$  is valid, iff,  $\mathcal{I}, y^{\mathcal{I}} \models \alpha$ , we obtain

$$\begin{aligned} \forall x^{\mathcal{I}} \forall y^{\mathcal{I}} \forall z_x (z_x \succeq x^{\mathcal{I}} \rightarrow \forall z_y (z_y \succeq y^{\mathcal{I}} \rightarrow \\ ((\mathcal{I}, z_y \models x : \forall R.\alpha) \wedge (\mathcal{I}, z_y \models x : \forall R.\alpha) \wedge ((z_x, z_y) \in R^{\mathcal{I}}) \rightarrow (\mathcal{I}, y \models \alpha)))) \end{aligned} \quad (3)$$

Consider the semantics of  $\exists R.\alpha$ :

$$(\exists R.\alpha)^{\mathcal{I}} =_{df} \{x \mid \exists y \in \mathcal{U}^{\mathcal{I}}. (x, y) \in R^{\mathcal{I}} \text{ and } y \in \alpha^{\mathcal{I}}\}$$

and the following rule:

$$\frac{\Delta \Rightarrow xRy \quad \Delta \Rightarrow y : \alpha}{\Delta \Rightarrow x : \exists R.\alpha} \exists\text{-r}$$

We can see that the premises of the rule entails the conclusion. The premises correspond to the following conditions:

$$\forall x^{\mathcal{I}} \forall y^{\mathcal{I}} \forall z_x (z_x \succeq x^{\mathcal{I}} \rightarrow \forall z_y (z_y \succeq y^{\mathcal{I}} \rightarrow ((z_x, z_y) \in R^{\mathcal{I}})))$$

and

$$\forall y^{\mathcal{I}} \forall z_y (z_y \succeq y^{\mathcal{I}} \rightarrow ((\mathcal{I}, z_y \models y : \alpha)))$$

Instantiating in both conditions  $z_y = y^{\mathcal{I}}$  and  $z_x = x^{\mathcal{I}}$ , this yields  $(x^{\mathcal{I}}, y^{\mathcal{I}}) \in R^{\mathcal{I}}$ , such that  $\mathcal{I}, y^{\mathcal{I}} \models \alpha$ , so  $\mathcal{I}, z_x \models x^{\mathcal{I}} : \exists R.\alpha$ . Thus,  $\exists\text{-r}$  is sound. The soundness of  $\exists\text{-l}$  is analogous to  $\forall\text{-l}$ .

Finally, it is worth noting that, for each rule, we can derive the soundness of its non-nominal version from the proof of soundness of its nominal version. For instance, the soundness of the nominal version of rule  $\sqcup\text{-l}$  depends on the diamond conditions F1 and F2. The soundness of its non-nominal version, is a consequence of the soundness of the nominal version.

The rules below have their soundness proved as a consequence of the following reasonings in first-order intuitionistic logic that are used for deriving the semantics of the conclusions from the semantics of the premises:

$$(\mathbf{p}\text{-}\exists): \forall x (A(x) \wedge B(x) \rightarrow C(x)) \models \forall x A(x) \wedge \exists x B(x) \rightarrow \exists x C(x);$$

$$(\mathbf{p}\text{-}\forall): (A(x) \models B(x)) \text{ implies } \forall y (R(y, x) \rightarrow A(x)) \models \forall y (R(y, x) \rightarrow B(x));$$

$$(\mathbf{p}\text{-}\mathbf{N}): \text{ if } A \models B \text{ then for every Kripke model } \mathcal{I} \text{ and world } x^{\mathcal{I}}, \text{ if } \mathcal{I}, x^{\mathcal{I}} \models A \text{ then } \mathcal{I}, x^{\mathcal{I}} \models B.$$

$$\frac{\Delta, \alpha \Rightarrow \beta}{\forall R.\Delta, \exists R.\alpha \Rightarrow \exists R.\beta} \mathbf{p}\text{-}\exists \quad \frac{\Delta \Rightarrow \alpha}{\forall R.\Delta \Rightarrow \forall R.\alpha} \mathbf{p}\text{-}\forall \quad \frac{\Delta \Rightarrow \delta}{x : \Delta \Rightarrow x : \delta} \mathbf{p}\text{-}\mathbf{N}$$

## 8. Conclusion

In this article, we shown how intuitionistic logic and Kelsen's jurisprudence can be used to express Chisholm paradox faithfully. A key fact in providing a logical model to this paradox is that laws/norms are not taken as propositions. For example, in the explanation above on building the model, if we turn back to deontic expression of laws, we will have that  $l_1$  is  $Op$  and  $l_2$  is  $O(p \rightarrow q)$ , but we cannot derive that  $l_3$  is  $O(q)$ .  $l_3$  is of course the meet ( $\sqcap$ ) between  $l_1$  and  $l_2$ , as a meet it is strongly connection to  $O(l_1) \wedge O(l_2) \leftrightarrow O(l_1 \wedge l_2)$ , which is a SDL valid formula. Thus,  $l_3$  is the norm  $O(l_1 \wedge l_2)$ , that is an obligation. However, now remembering what norms  $l_1$  and  $l_2$  are in this particular case,  $l_3$  is the meeting  $O(p) \wedge O(p \rightarrow q)$  that it is  $O(p \wedge (p \rightarrow q))$ . This conclusion, however, does not entail that in  $l_3$  can be identified with  $O(q)$ , since our implication is the intuitionistic implication. This very last aspect of joining Kelsen jurisprudence and iALC also helps to avoid other deontic paradoxes.

Jorgensen's Dilemma (Jorgensen, 1937) offers a question, in fact, a dilemma, whether there is, in fact, any deontic logic. The question follows this path: 1) Norms/laws deal with evaluative sentences; 2) Evaluative sentences are not the kind of sentence that can be true or false; 3) Thus, how there is a logic of evaluative sentences? 4) Logic has as goal to define what can be drawn from whatever, and; 5) A sentence follows from a set of sentences on a basis of the relationship between the truth of the sentences in question. Thus, there is no deontic logic. What we have shown in this article, is that deontic logic is possible by considering the logic of norms as a logic on norms, instead. This reading is just what we do in legal ontologies.

We have to touch some aspects that are very well-known in the deontic approach. One is the deontic concept of *permission*. This case is modeled by observing that in a society regulated by law, permission is nothing more than an obligation of the State. The State promulgates what is allowed. Concerning prohibitions, the foundation is analogous. However, some subtle and theoretical problems may arise if one wants to recover the definition of forbidden ( $F$ ) regarding the very well-known duality  $F(p) \equiv O(\neg p)$ . This discussion will be the subject of another article.

On the more practical side, we have been experimenting with applications of iALC aiming to develop a methodology for law formalization. The OAB (Ordem dos Advogados do Brasil, the Order of Attorneys of Brazil) Exam can be considered an excellent choice for analyzing the performance of a system with the goal to reason in the legal world. Similarly to the US Bar Exam, it evaluates whether the candidates have the aptitude to practice the law. We have conducted three experiments in question answering on the OAB national exams (Delfino, Cuconato, Haeusler, & Rademaker, 2017; Alkmim, Haeusler, & Rademaker, 2018). In these articles we explain the basic structure of the OAB Exam followed by the formal justification of the corrected answers of some questions. In the future, we intend to make these results as general and algorithmic as possible, aided by a reasoner for iALC and combined with natural language processing techniques.

## Bibliography

- Alchourron, C. E., & Martino, A. (1990). Logic without truth. *Ratio Juris*, 3(1), 46–67.
- Alkmim, B., Haeusler, E. H., & Rademaker, A. (2018). Utilizing ialc to formalize the brazilian oab exam. In *Proceedings from the explainable ai in law workshop (xaila)*. CEUR-WS. (to appear)
- Bellin, G., de Paiva, V., & Ritter, E. (2001). Extended curry-howard correspondence for a basic constructive modal logic. In *Procs of methods for the modalities*.
- Braüner, T., & de Paiva, V. (2006). Intuitionistic hybrid logic. *JAL*, 4(3), 231–255.
- Buss, S. R. (1998). *Handbook of proof theory*. Elsevier.
- Clément, I. (2008). *Proof theoretical foundations for constructive description logic* (Unpublished master’s thesis). School of Computer Science. McGill University, Montreal.
- Delfino, P., Cuconato, B., Haeusler, E. H., & Rademaker, A. (2017). Passing the brazilian oab exam: Data preparation and some experiments. In A. Wyner & G. Casini (Eds.), *Legal knowledge and information systems* (Vol. 302). (The 30th International Conference on Legal Knowledge and Information Systems (JURIX 2017). Expanded version at <https://arxiv.org/abs/1712.05128>)
- de Paiva, V. (2006). Constructive description logics: what, why and how. In *Context representation and reasoning*.
- de Paiva, V., Hausler, E. H., & Rademaker, A. (2010). Constructive description logic: Hybrid-style. In *Proceedings of hybrid logic and applications (HyLo 2010)*.
- Fischer-Servi, G. (1984). Axiomatizations for some intuitionistic modal logics. *Rendiconti del Seminario Matematico Università e Politecnico di Torino*, 42.
- Girard, J., Taylor, P., & Lafont, Y. (1989). *Proofs and types*. Cambridge University Press.
- Haeusler, E. H., de Paiva, V., & Rademaker, A. (2010). Intuitionistic logic and legal ontologies. In *Proc. jurix 2010* (pp. 155–158). IOS Press.
- Haeusler, E. H., de Paiva, V., & Rademaker, A. (2010). Using intuitionistic logic as a basis for legal ontologies. In *Proceedings of the 4th workshop on legal ontologies and artificial intelligence techniques* (pp. 69–76). Fiesole, Florence, Italy: European University Institute.
- Haeusler, E. H., de Paiva, V., & Rademaker, A. (2011). Intuitionistic description logic and legal reasoning. In *Proceedings of international workshop data, logic and inconsistency with dexa 2011*.
- Haeusler, E. H., Paiva, V. D., & Rademaker, A. (2010). Using intuitionistic logic as a basis for legal ontologies. *Informatica e Diritto*, XIX(1-2), 289–298.
- Hughes, G. (1971). Validity and the basic norm. *California Law Review*, 59(3), 695–715.
- Jorgensen, J. (1937). Imperatives and logic. *Erkenntnis*, 7, 288–296.
- Kelsen, H. (1946). *General theory of law and state*. Harvard University Press. (Translation from german by Wedberg, A. and Kraus, W.H.)
- Kelsen, H. (1967). *Pure theory of law* [Book]. Berkeley : University of California Press. (Translated from *Reine Rechtslehre*, second edition, 1960)

- Kelsen, H. (1991). *General theory of norms*. USA: Oxford Univ. Press.
- M.E. Coniglio, N. P. (2009). A paraconsistent approach to chisholm's paradox. *Principia*, 13(3), 299–326.
- Mendler, M., & Scheele, S. (2010). Towards constructive DL for abstraction and refinement. *JAR*, 44(3), 207–243. (Proc. 21st International DL Workshop)
- Negri, S., & Von Plato, J. (2008). *Structural proof theory*. Cambridge University Press.
- Plotkin, G., & Stirling, C. (1986). A framework for intuitionistic modal logics: extended abstract. In *Proceedings of the 1986 conference on theoretical aspects of reasoning about knowledge* (pp. 399–406). San Francisco, CA, USA: Morgan Kaufmann Publishers Inc. Retrieved from <http://dl.acm.org/citation.cfm?id=1029786.1029823>
- Rademaker, A. (2012). *A proof theory for description logics*. Springer.
- Ranalter, K. (2010). Embedding constructive  $k$  into intuitionistic  $k$ . *Electr. Notes Theor. Comput. Sci.*, 262, 205–219.
- Raz, J. (1972). Legal principles and the limits of law. *Yale Law Journal*, 81, 823–854.
- Simpson, A. (1993). *The proof theory and semantics of intuitionistic modal logic* (Unpublished doctoral dissertation). University of Edinburgh.
- Takeuti, G. (2013). *Proof theory*. Courier Dover Publications.
- Valente, A. (1995). *Legal knowledge engineering: A modelling approach*. IOS Press. (Amsterdam)



# ALGUNS COMENTÁRIOS SOBRE A QUESTÃO DA IDENTIDADE ENTRE LÓGICAS

Hugo Luiz Mariano<sup>1</sup>

## 1. Introdução

O objetivo destas notas, motivadas por palestra proferida no “*Lógica no Avião*” em setembro de 2016, não é realizar uma discussão conceitual abrangente ou aprofundada sobre o tema da identidade entre entes matemáticos, especificamente no contexto das lógicas, mas apresentar ao leitor interessado um pequeno recorte sobre o tema, focado em alguns encaminhamentos matemáticos, realizados em trabalhos de alguns lógicos, alguns dos quais com atuação ou parcerias acadêmicas no Brasil. Não há qualquer originalidade em abordar o tema sobre identidade/equivalência entre sistemas lógicos, sendo este um assunto recorrente na comunidade dos lógicos; reflexões desta natureza podem ser encontradas, por exemplo, em [5], [10] e [16]. Registramos aqui também que partes do texto [38] foram utilizadas para compor estas presentes notas.

Um primeiro esclarecimento já caberia aqui, originado pela distinção que ocorre na prática entre Lógica como ramo do conhecimento útil para os Fundamentos da Matemática, e a Lógica enquanto área da Matemática com diversas aplicações em outras áreas da própria Matemática (como em Teoria de Modelos). Se por um lado, na primeira concepção da Lógica, é adequado proceder tão mais finitariamente enquanto possível<sup>2</sup>, já no segundo caso, como em Teoria dos Modelos, assumi-se que uma meta-teoria dos conjuntos muito mais rica contendo, por exemplo, ZFC (pelo menos). Nestas notas, por encararem lógicas enquanto objetos matemáticos, vão na segundo direção<sup>3</sup>.

Ilustrando a questão da identidade entre objetos matemáticos:

Consideremos um estudante que entra em contato com a noção matemática denominada *grupo*, i.e. uma estrutura matemática  $(G, \cdot, 1)$ , dada por um conjunto (não vazio)  $G$ , munido de alguns dados: uma operação binária de “multiplicação”  $(\cdot)$  que é associativa, e que possui um elemento distinguido  $(1)$  que é elemento neutro para esta operação binária, e tal que, além disso, todo elemento admite um (único) inverso (para todo  $g \in G$ , existe  $h \in G$  tal que  $g.h = 1 = h.g$ ), ou seja, a estrutura poderia estar (ou estaria) munida de uma operação unária extra  $((\ )^{-1})$  que descreve os inversos multiplicativos de cada elemento. Este estudante não teria dificuldade de aceitar que ambas as apresentações da noção de grupo (com ou sem uma operação unária na linguagem) têm “o mesmo conteúdo matemático”, apesar de uma estar dada em linguagem mais completa e satisfazer axiomas mais simples (todos universais) e a outra estar dada em uma linguagem mais restrita e

---

<sup>1</sup> Instituto de Matemática e Estatística, Universidade de São Paulo. Email: [hugomar@ime.usp.br](mailto:hugomar@ime.usp.br).

<sup>2</sup> Assumindo portanto que na meta-teoria dispomos apenas de uma fraca (meta)teoria de conjuntos (se assim a desejarmos formalizar) que permite ao menos fazermos definições por recursão e provas por indução.

<sup>3</sup> A semântica usual da Lógica de predicados (clássica, finitária) de primeira ordem ou de ordem superior e a própria formulação do que são lógicas infinitárias só podem ser postas no contexto de teorias robustas de conjuntos como ZFC.

necessitar ser descrita por axiomas mais complexos ( $\forall\exists$ ). Tecnicamente isto fica respaldado na Lógica de primeira ordem através do “Teorema de Extensão por Definições”, cuja formulação pode ser encontrada nos bons livros de Lógica Matemática, e.g. no livro do Shoenfield. Vale acrescentar que mesmo esta questão da indistinguibilidade entre teorias matemáticas de primeira ordem admite complexidades e variações não triviais, que envolvem alterações “significativas” de linguagens envolvidas, mas não cabe nestas notas nos determos no detalhamento destas questões<sup>4</sup>.

Entramos agora na questão central destas notas, a indistinguibilidade de lógicas, ilustrando-a inicialmente com pontos específicos:

Via de regra, o primeiro contato de um estudante com lógica simbólica ocorre através de alguma descrição da Lógica Proposicional Clássica e a interpretação dos conectivos básicos ( $\wedge$ ,  $\vee$ ,  $\neg$ ,  $\rightarrow$ ) por meio de tabelas de verdade a dois valores,  $\{0, 1\}$ , que se estende às interpretações das correspondentes fórmulas proposicionais determinadas por este conjunto de conectivos e um conjunto infinito enumerável de variáveis proposicionais. A este aspecto inicial semântico costuma seguir uma descrição axiomática desta lógica no estilo de Hilbert, i.e. por meio de axiomas e regras de inferência formulados em “parte representativa” desta mesma linguagem. E segue-se usualmente com o estabelecimento de importantes (meta)teoremas, como os metateoremas da Correção e da Completude<sup>5</sup>, que “identificam” essas apresentações semântica e axiomática.

Coloca-se naturalmente neste quadro, a questão semântica dos subconjuntos de conectivos que “determinam a mesma informação” (e.g.  $\{\neg, \rightarrow\}$ ,  $\{\neg, \vee\}$ ,  $\{\neg, \wedge\}$ ) ou mesmo quando esta invariância de conteúdo ocorre através de conjuntos que contam com a presença de conectivos alternativos (como o formado apenas pelo símbolo de Sheffer). Há também uma diversidade de descrições axiomáticas alternativas, formuladas em diversas possibilidades de linguagem. Haveria ainda outras alternativas prova-teóricas de capturar o conteúdo dessa lógica além da apresentação axiomática à lá Hilbert, e.g. cálculo de sequentes, dedução natural, tableaux,... Ainda assim, dizemos que estamos nos referindo a mesma lógica proposicional clássica! Mas afinal, o que é, concretamente, a Lógica Proposicional Clássica?([11])

Claro que o problema não se restringe a questão pontual das apresentações da lógica proposicional clássica. A lógica proposicional intuicionista, formulada na linguagem ( $\wedge$ ,  $\vee$ ,  $\neg$ ,  $\rightarrow$ ) é reconhecida como uma *sublógica* da lógica clássica (i.e.  $\Gamma \vdash_i \phi \Rightarrow \Gamma \vdash_c \phi$ ) e sendo de fato uma sublógica *própria*, algo testemunhado, por exemplo, pela ausência do axioma (ou teorema) do terceiro excluído. Reversamente, a toda demonstração clássica está naturalmente associada uma demonstração intuicionista através da “tradução de Glivenko”:  $\Gamma \vdash_c \phi \Rightarrow \neg\neg\Gamma \vdash_i \neg\neg\phi$ . Nos anos de 1930, Gödel verificou que, diferentemente da lógica clássica, a lógica intuicionista não pode ser caracterizada por uma semântica finita; por outro lado, provou também que o fragmento ( $\neg$ ,  $\wedge$ ) da lógica intuicionista “coincide” com a lógica clássica. Assim o problema da identidade de lógicas sugere também

<sup>4</sup>Outras noções razoáveis de equivalência entre teorias matemáticas são dadas pelos conceitos de bi-interpretação e interpretação mútua entre teorias: para estas, variantes e outras questões envolvendo aspectos ontológicos de teorias como teorias aritméticas e teorias de conjuntos e classes veja, por exemplo, [3] e [25]. Outras noções de identidades de teorias ainda são possíveis e começaram a ser mais exploradas recentemente, como a noção (categorial) de teorias *Morita-equivalentes*, que tem sido desenvolvida principalmente por Olivia Caramello.

<sup>5</sup> Este último pode ser estabelecido por pelo menos dois métodos diferentes: há uma abordagem mais efetiva, denominada a prova Kálmár, e outra de cunho algébrico, dada por meio da utilização da álgebra de Boole naturalmente associada à Lógica Clássica, denominada sua Álgebra de Lindenbaum.

a importância e utilidade em se estabelecer critérios mais amplos de comparação entre lógicas.

Tendo em vista o comentário acima, nosso ponto de partida sobre o tema geral da identidade de lógicas é o (aparentemente) mais simples possível, o critério de Leibniz de identidade:

*“Objetos  $A, A'$  são idênticos precisamente quando são indiscerníveis, i.e.  $A$  e  $A'$  compartilham as mesmas propriedades.”*

Algumas questões surgem imediatamente: O que seria razoável aceitarmos como sendo “propriedades”? Quais seleções de propriedades poderiam estar em jogo? Como nos referirmos aos “dois” objetos  $A, A'$  que viriam a ser idênticos?

Destas questões, a terceira é mais simples de ser considerada: basta adotarmos o encaminhamento standard de distinguir linguagem objeto e metalinguagem.

Já sobre as duas primeiras questões a situação é muito mais delicada. Um encaminhamento matematicamente simples e natural não seria listar explicitamente o que se pretende aceitar como “propriedades”, mas tê-las determinado implicitamente. Mais concretamente: a noção matemática de *categoria* cumpre bem este papel, consistindo de seleções de uma classe de objetos que são conectados por meio de uma classe selecionada de morfismos, que podem ser compostos de modo a satisfazer leis muito simples, veja por exemplo [32]. O linguajar da Teoria das Categorias fornece imediatamente uma versão simples e razoável da noção de identidade estrutural, dada pelos isomorfismos (= morfismos invertíveis) entre objetos da categoria. Isto já nos leva aos passos seguintes destas notas: considerar categorias cujos objetos “são lógicas”. Cabe aqui dizer que, além disso, a Teoria das Categorias também permite estabelecer critérios de comparações de “seleções implícitas de propriedades”: por exemplo, as duas apresentações da noção de grupo mencionadas anteriormente são equivalentes pois essas determinam categorias isomorfas.

## 2. Motivação principal

A definição e o estudo sistemático de categorias cujos objetos são lógicas (geralmente proposicionais e com conectivos finitários) foi motivado pelo fenômeno do surgimento de diversos processos de combinações de lógicas que ocorreu na década de 1990, a saber os processos de: produto, fusão, temporalização, sincronização, parametrização e fibrilação (irrestrita e restrita), veja por exemplo [14]. Em geral, estes processos foram definidos em contextos de lógicas particulares (e.g. para lógicas modais) e buscava-se determinar condições que garantissem que fossem preservadas, sob o processo de combinação, certas propriedades meta-lógicas interessantes das lógicas constituintes como as propriedades de: Correção, Completude (ou Completude fraca), Interpolação de Craig, Decidibilidade, etc.

Esses processos de combinação de lógicas se apresentam em aspectos duais: análise de lógicas e síntese de lógicas. Em um processo de combinação, representado aqui por  $\otimes$ , a relação entre uma lógica  $L$  e uma família de lógicas  $\{L_i : i \in I\}$ , através da “equação”  $L \cong \otimes_{i \in I} L_i$ , pode ser encarada de duas formas:

- (i) como um processo de decomposição ou análise de lógicas: se a lógica  $L$  já foi previamente definida ou considerada, mas está em estudo, e as lógicas da família  $\{L_i : i \in I\}$  são bem conhecidas, frequentemente tratando-se de fragmentos mais simples de  $L$ ;
- (ii) como um processo de composição ou síntese de lógicas: se as lógicas da família  $\{L_i : i \in I\}$  são conhecidas e pretende-se definir (e estudar as meta-propriedades) da

lógica combinada  $L \cong \bigotimes_{i \in I} L_i$ ; ou se as lógicas  $\{L\} \cup \{L_i : i \in I\}$  são definidas e conhecidas com a equação significando que foi identificada uma relação entre essas.

Vamos descrever brevemente exemplos típicos de cada um desses aspectos. No início dos anos de 1990, W. Carnielli introduziu o conceito de Semântica de Traduções Possíveis, que baseia-se na idéia de descrever/decompor/analisar uma lógica por meio de uma família de traduções (globalmente conservativa) em lógicas supostamente mais simples ou mais conhecidas que a original ([13]). Dualmente, D. Gabbay, em meados da década de 1990, considerou (originalmente no contexto das lógicas modais e semânticas de Kripke) a noção de fibrilação de lógicas, cujo objetivo seria combinar duas lógicas  $L_1$  e  $L_2$  de tal forma a obter "o menor sistema lógico na linguagem combinada que é uma extensão conservativa de ambas as lógicas  $L_1$  e  $L_2$ " ([27]).

Naturalmente, o ponto de partida para tratar categorialmente a questão da combinação de lógicas passa por dois processos de escolha: o de representar os sistemas lógicos que constituem os objetos da categoria em questão e o de descrever os tipos de morfismos entre esses. Dentro do processo de representação dos sistemas lógicos, fixada uma linguagem proposicional (assinatura) e a noção de fórmula correspondente, existem três possibilidades a serem consideradas, com várias alternativas em cada um dessas: (i) a representação sintática das lógicas, dentro do contexto da teoria da prova (sistemas axiomáticos à lá Hilbert, dedução natural, cálculo de seqüentes, tableaux, etc); (ii) a representação semântica (valorações, semântica de Kripke, matrizes lógicas, etc); (iii) a representação simultânea ou mista: "lógica = demonstrabilidade + semântica". Fixada a classe de objetos da categoria passa-se à questão da escolha de morfismos conectando esses objetos.

### 3. Uma escolha simples

Frequentemente a maneira mais simples e "concreta" de representar os sistemas lógicos proposicionais é a representação prova-teórica de lógicas como um par  $(\Sigma, \vdash)$ , onde:

- $\Sigma = (\Sigma_n)_{n \in \mathbb{N}}$  é uma assinatura proposicional e finitária;
- $F_\Sigma(X)$  é a  $\Sigma$ -álgebra completamente livre sobre um conjunto fixado  $X$  de "variáveis proposicionais" que é infinito e enumerável;
- $\vdash \subseteq P(F_\Sigma(X)) \times F_\Sigma(X)$  é relação de consequência "tarskiana" ([31]) associada a um operador de fecho (das "teorias")  $\mathcal{T} : P(F_\Sigma(X)) \rightarrow P(F_\Sigma(X))$  que é finitário e estrutural (i.e. é compatível com todos os  $\Sigma$ -homomorfismos de substituição  $h : F_\Sigma(X) \rightarrow F_\Sigma(X)$ ): estes operadores sempre são provenientes de algum sistema hilbertiano<sup>6</sup>.

As conexões mais simples e naturais entre os sistemas lógicos como descritos acima são dadas por (alguma variante da noção de) morfismos de tradução (ou interpretação), i.e. certas flechas associadas a funções, definidas nos conjuntos de fórmulas da lógica fonte na lógica alvo, que são "contínuas", i.e. que preservam as relações de consequência envolvidas.

Retornado à motivação principal para considerar categorias de lógicas, observamos que no final dos anos de 1990, o Grupo de Lógica do IST de Lisboa-Portugal e, posteriormente o Grupo de Lógica Teórica e Aplicada do CLE-UNICAMP, consideraram sistematicamente a perspectiva categorial do fenômeno de (certos tipos de) combinações

<sup>6</sup> A assinatura e os axiomas e regras de inferência do sistema hilbertiano providenciam uma apresentação da lógica em questão por meio de "geradores e relações" de forma análoga ao que ocorre para as estruturas algébrica dadas por alguma teoria equacional (como as teoria dos grupos, dos anéis, dos reticulados, etc) podem ser apresentadas por geradores e relações.

de lógicas, focalizando atenção em algumas construções categoriais em certas categorias de lógicas, onde procuravam determinar as condições necessárias para que certas meta-propriedades fossem conservadas pelas construções. Enquanto o processo analítico que ocorre na Semântica de Traduções Possíveis produz uma tradução conservativa da lógica em estudo em um produto (ou produto fraco) de lógicas mais simples, por outro lado a fibrilação "irrestrita", i.e. sem compartilhamento de conectivos das lógicas constituintes, é descrita pela construção categorial dual de produto, o coproduto, e a noção co-irmã a de fibrilação "restrita", i.e. com compartilhamento de conectivos das lógicas constituintes, é descrito pela construção categorial dual de produto fibrado (ou pullback), a soma amalgamada (ou pushout) ([43], [15]).

A abordagem categorial de noções como a de fibrilação é relevante porque, além desta exigir que os objetos de estudo e suas interrelações sejam totalmente precisos, a caracterização da fibrilação como uma construção universal (como coproduto ou como soma amalgamada) em uma dada escolha de categoria de lógicas, permite a *definição*, através das mesmas propriedades universais, da noção de fibrilação de lógicas em outras categorias que capturam outros aspectos dos sistemas lógicos: desta forma, foram propostas novas categorias de lógicas que apresentam tratamentos de dois problemas antípodos que ocorreram em certas fibrilações nas primeiras categorias de lógicas: o problema do *colapso* ([42], ([18])) e o do *anti-colapso* ([17]).

Como exposto acima, a corrente principal do estudo das categorias de lógicas focalizou sua atenção em algumas construções categoriais em certas categorias de lógicas, procurando determinar condições suficientes para que certas meta-propriedades sejam conservadas pelas construções. Posteriormente os aspectos "globais" das categorias de lógicas, que garantem por exemplo a abundância (ou escassez) de construções - algo fundamental para os processos de combinações de lógicas - estabeleceu-se como projeto de investigação complementar ao desta corrente principal. Dentre os trabalhos nesta linha, e que têm em comum essa natureza mais simples de apresentar os sistemas lógicos enquanto objetos, listamos: [22], [1], [16], [33], [5].

Em [1], além da representação simples e concreta dos sistemas lógicos por meio de um par formado por uma assinatura e uma relação de consequência tarskiana, tem-se a noção mais simples (porém restrita) e natural possível de morfismo entre assinaturas e de morfismo de tradução induzido entre lógicas: um morfismo (dito estrito) de assinaturas  $f : \Sigma \rightarrow \Sigma'$  é simplesmente uma sequência de funções que, a cada conectivo  $n$ -ário da assinatura fonte,  $\Sigma$ , atribui um conectivo  $n$ -ário da assinatura alvo,  $\Sigma'$ ; este induziria uma função  $\hat{f} : F_{\Sigma}(X) \rightarrow F_{\Sigma'}(X)$  que, caso preservasse as relações de consequência (i.e.  $\Gamma \vdash \phi \Rightarrow \hat{f}[\Gamma] \vdash' \hat{f}(\phi)$ ), determinaria um morfismo (estrito) de lógicas  $f : (\Sigma, \vdash) \rightarrow (\Sigma', \vdash')$ .  $\mathcal{S}_s$  e  $\mathcal{L}_s$  denotam as respectivas categorias de assinaturas e de lógicas obtidas através de definições naturais de composição de morfismos e de morfismos de identidades.  $\mathcal{S}_s$  e  $\mathcal{L}_s$  têm boas propriedades categoriais pois são completas, co-completas e (finitamente) acessíveis (veja [6] para este conceito): estes resultados constituem o alicerce para outros de resultados apresentados no trabalho em questão.

Como já foi mencionado, uma apresentação de uma lógica como colimite de outras pode ser entendido como uma *decomposição* da lógica em estudo por meio de outras lógicas (conhecidas e/ou bem comportadas), possivelmente auxiliando no entendimento da lógica mais complexa. Em [1] observou-se que a situação onde uma lógica  $L$  é completamente determinada pelas traduções de outras lógicas  $L_i$  em  $L$  pode ser percebida como uma

“cobertura” de  $L$  pelas lógicas  $L_i$ ; nesse trabalho transformou-se esta intuição inicial em afirmações matemáticas ao fixar uma categoria de sistemas lógicos e fornecer algumas definições rigorosas da noção de cobertura. Além disso, cada noção de cobertura tomada induz uma *topologia de Grothendieck* na (essencialmente pequena) subcategoria plena  $(\mathcal{L}_s)_{fp} \subseteq \mathcal{L}_s$  formada pelas lógicas finitamente presentáveis. Com as noções de *feixe* assim obtidas foi possível relacionar lógicas (= colimites filtrantes de lógicas finitamente presentáveis) e feixes por intermédio da categoria  $\mathbf{Set}^{((\mathcal{L}_s)_{fp})^{op}}$  dos préfeixes (= colimites de funtores contravariantes *hom* de lógicas finitamente presentáveis).

Prosseguindo neste ponto de vista global da lógica, estudou-se em [1] uma categoria cujos objetos são as lógicas algebrizáveis no senso de Blok-Pigozzi<sup>7</sup> (veja [12]). Demonstrou-se que a categoria  $\mathcal{A}_s$ , formada pelas lógicas Blok-Pigozzi algebrizáveis com morfismos tomados como os morfismos estritos de lógicas e que também preservam pares algebrizadores (i.e. as equações definidoras e as fórmulas de equivalência da lógica algebrizável), é uma categoria que possui todos os colimites filtrantes, constituindo também uma categoria finitamente acessível. Porém destacamos que esta não é completa, pois não possui limites de diagramas “não limitados”, nem cocompleta, pois não possui objeto inicial (nem objeto inicial fraco em geral).

Como mencionamos, foram apresentados em [1] os *primeiros passos* para se adotar a perspectiva global em Lógica (e se obter informações relevantes a partir desta abordagem) ao tomar escolhas simples e naturais de categorias de assinaturas e de sistemas lógicos. Contudo, na prática, essas escolhas se revelam muito restritivas: tratando de forma inadequada o problema da identidade de sistemas lógicos. Por exemplo, duas apresentações da lógica proposicional clássica que tomam como conjuntos de conectivos primitivos (i.e. assinaturas)  $\{\neg, \rightarrow\}$  e  $\{\neg, \vee\}$  não admitem nenhum morfismo estrito lógico entre elas (desde que esse precisaria levar  $\rightarrow$  em  $\vee$ ) enquanto seria natural esperar que essas apresentações fossem isomorfas. Para ir na direção de remediar esse defeito, considerou-se em [22] tomar uma noção mais flexível de morfismo de assinaturas que permitisse levar em conta substituições nas álgebras de fórmulas, admitindo assim mapear o conectivo binário primitivo  $\rightarrow$  no conectivo binário “derivado”  $(\neg(x_0) \vee x_1)$  (i.e. uma fórmula em duas variáveis) e onde o conectivo unário  $\neg$  seria mapeado em uma fórmula em uma variável  $\neg x_0$ ; na direção reversa, também teríamos outro morfismo flexível dado por  $\vee \mapsto (\neg(x_0) \rightarrow x_1)$  e  $\neg \mapsto \neg x_0$ . Assim, mapeando-se em geral os conectivos  $n$ -ários da assinatura fonte do morfismo (dito flexível) em conectivos “derivados”  $n$ -ários da assinatura alvo (i.e., uma fórmula da linguagem alvo onde aparecem exatamente as  $n$ -primeiras variáveis) pode-se obter uma categoria de assinaturas e morfismos flexíveis,  $\mathcal{S}_f$ , e sua correspondente categoria de lógicas e morfismos flexíveis de tradução  $\mathcal{L}_f$ . Contudo, analisando complexidade de fórmulas, é fácil verificar que estes mapeamentos *não* produziram isomorfismos inversos de assinaturas ou de lógicas. Além disso, a categoria de assinaturas resultante é bastante

<sup>7</sup> Ao passo que o método de algebrização introduzido por Lindenbaum e Tarski de atribuir uma álgebra a uma dada lógica através de uma construção explícita, de modo a fazer um interplay interessante entre a apresentação prova-teórica e a contraparte semântica da lógica em questão, a teoria das lógicas algebrizáveis desenvolvida por Blok e Pigozzi - que, em certo sentido, generaliza aspectos do método anterior - descreve conceitualmente a forma de corresponder sintaxe e semântica (algébrica) em lógicas assim ditas algebrizáveis, através de conjuntos finitos de “fórmulas de equivalência” (fórmulas em duas variáveis) e de um conjunto finito de “equações definidoras” (i.e. pares de fórmulas em uma variável), e tal que ambos os dados detêm uma relação estreita com a tal lógica dita algebrizável; em particular existe uma única quasivariabilidade de álgebras na mesma assinatura da lógica, denominada semântica algébrica equivalente da lógica, na qual se produz uma forma especial de teoremas de correção e completude.

pobre em termos de construções: e.g. ela não possui pullbacks nem colimites filtrantes. Em [16] foi introduzida uma noção variante de comparação mútua de lógicas através de morfismos flexíveis,  $(\Sigma, \vdash) \xrightleftharpoons[h]{h'} (\Sigma', \vdash')$ , denominada *equipolência*, que pode ser descrita grosso modo dizendo que os morfismos flexíveis  $h, h'$  são “inversos a menos de interdemonstrabilidade”: rapidamente verifica-se que as duas apresentações da lógica clássica dadas acima seriam “equipolentes”.

Posteriormente surgiram duas abordagens distintas ([33], [5]), porém em certa medida equivalentes, que refinam os trabalhos mencionados anteriormente ao para lidar simultaneamente, e de forma efetiva, com o problema da identidade de lógicas e a abundância de construções categoriais:

Em [33], foram consideradas categorias de lógicas satisfazendo *simultaneamente* certos requisitos naturais, tais como:

- (i) Representar a maioria dos sistemas lógicos usuais;
- (ii) Fornecer um tratamento satisfatório do *problema de identidade* de sistemas lógicos;
- (iii) Possuir boas propriedades categoriais (por exemplo, se elas são categorias completas e/ou cocompletas, se elas são categorias acessíveis);
- (iv) Permitir uma noção natural de sistema lógico *algebrizável* (como no conceito de lógica algebrizável segundo Blok-Pigozzi ([12]) ou de proto-algebrizabilidade de Czelakowski ([19])).

Inicialmente, foram analisadas as duas categorias de assinaturas (finitárias) consideradas na literatura:  $\mathcal{S}_s$  tem morfismos mais simples mas muito estritos e melhores propriedades categoriais;  $\mathcal{S}_f$  tem morfismos mais flexíveis mas é categorialmente mais pobre. Definiu-se um par de funtores adjuntos entre essas categorias, mostrando que a mônada ou tripla  $\mathcal{T} = (T, \mu, \eta)$  sobre  $\mathcal{S}_s$ , canonicamente associada a esta adjunção, é tal que  $T$  preserva colimites filtrantes, reflete isomorfismos e, principalmente, que  $Kleisli(\mathcal{T}) = \mathcal{S}_f$ , de onde se deriva algumas informações adicionais sobre a categoria  $\mathcal{S}_f$ : por exemplo, esta tem todos coprodutos. Segue-se uma análise da categoria de lógicas  $\mathcal{L}_f$  associada a categoria de assinaturas  $\mathcal{S}_f$ : é verificado que a única obstrução para a existência de um “tipo” de limite/colimite em  $\mathcal{L}_f$  é a não existência deste “tipo” de limite/colimite em  $\mathcal{S}_f$ : isto já permite concluir que existem em geral (i.e., mesmo no caso infinitário) todas as fibrilações irrestritas e fibrilações restritas quando os  $\mathcal{S}_f$ -morfismos de fonte comum são regulares, i.e. são imagens de  $\mathcal{S}_s$ -morfismos. Define-se a categoria quociente  $\mathcal{Q}_f := \mathcal{L}_f / \dashv\vdash$ , que identifica morfismos flexíveis paralelos entre lógicas cujas imagens de cada fórmula da lógica fonte são inter-demonstráveis na lógica alvo, verifica-se que os  $\mathcal{Q}_f$ -isomorfismos coincidem com as equipolências de [16] e que, portanto, na categoria quociente  $\mathcal{Q}_f$  diversas apresentações<sup>8</sup> da lógica clássica são, de fato, *isomorfas*. Considera-se uma subcategoria plena  $\mathcal{L}_f^c \hookrightarrow \mathcal{L}_f$ , constituída pelas lógicas ditas “congruenciais” e verifica-se que esta é uma subcategoria reflexiva: logo a toda lógica está canonicamente associada uma lógica congruencial optimal. Finalmente, na correspondente categoria quociente  $\mathcal{Q}_f^c := \mathcal{L}_f^c / \dashv\vdash$ : (i) resolve-se o “problema da identidade” para lógica clássica em termos de  $\mathcal{Q}_f^c$ -isomorfismos; (ii) a toda lógica esta associada optimalmente um objeto de  $\mathcal{Q}_f^c$ ; (iii)  $\mathcal{Q}_f^c$  possui boas propriedades categoriais: é completa, cocompleta e finitamente acessível (i.e. toda lógica congruencial é colimite filtrante, em  $\mathcal{Q}_f^c$ , de lógicas “finitamente presentáveis”); (iv) existem uma escolha natural de noção de lógica “Lindenbaum” algebrizável, e uma escolha de morfismo

<sup>8</sup> Que não possuem símbolos de constante lógicas como  $\top$  e  $\perp$ .

entre essas, de modo a determinar uma categoria  $Lind(\mathcal{A}_f)$  tal que  $Lind(\mathcal{A}_f) = \mathcal{A}_f \cap \mathcal{L}_f^c$  e assim obtendo uma boa subcategoria  $\mathcal{Q}(Lind(\mathcal{A}_f)) \subseteq \mathcal{Q}_f^c$ .

Em [5], leva-se a campo a teoria abstrata da homotopia aplicada ao contexto das categorias de lógicas. Desta forma tem-se uma “categoria relativa” formada pelo par  $(\mathcal{L}_f, \mathcal{W})$ , onde  $\mathcal{W}$  é a classe dos morfismos flexíveis que são “equivalências fracas”, i.e. as traduções conservativas que são “morfismos densos”. Nesta teoria relativa de categorias estão disponíveis as noções de limites e colimites homotópicos, que grosso modo constituem as melhores aproximações de uma construção de limite/colimite que preservam as equivalências fracas. É estabelecido que  $(\mathcal{L}_f, \mathcal{W})$  possui todos os limites e colimites homotópicos. Verifica-se que as equivalências fracas coincidem com as equipolências e que, portanto, a correspondente categoria homotópica (de frações),  $\mathcal{L}_f[\mathcal{W}^{-1}]$ , é equivalente a categoria quociente  $\mathcal{Q}_f = \mathcal{L}_f/\dashv\vdash$ , situação análoga a que ocorre no contexto das categorias dos espaços topológicos e equivalências de homotopia. Desta forma, esta abordagem não permite apenas uma abundância de construções (homotópicas) e um bom tratamento da questão da identidade de lógicas, mas ela também permite que muitas intuições da teoria da homotopia, desenvolvidas no contexto dos espaços topológicos, sejam transpostas ao contexto das categorias de lógicas.

Finalizamos esta seção observando que mesmo as noções ditas flexíveis de morfismos de traduções de lógicas não englobam a tradução de Glivenko entre a Lógica Clássica e a Lógica Intuicionista. Este ponto sugere que também seria natural considerar ampliações do cenário utilizado nesta seção, tanto na questão da representação de flechas ditas de tradução quanto na representação dos sistemas lógicos como objetos. Desenvolveremos isto na sequência.

#### 4. Lógicas abstratas e representações

Nesta seção esboçaremos, com menor nível de detalhes que na seção anterior, alguns desenvolvimentos da apresentação de lógicas e suas conexões que são mais abstratos e/ou dados por meios menos diretos de representação.

Iniciamos observando que, ao invés de se considerar apenas os sistemas lógicos “concretos” como na seção anterior, i.e., aqueles dados por um par  $(\Sigma, \vdash)$  onde  $\vdash$  está associado a um certo operador de fecho definido na  $\Sigma$ -álgebra de fórmulas  $F_\Sigma(X)$ , pode ser considerada uma ampliação natural desta idéia, considerando agora uma lógica como dada por um par  $(S, \vdash)$  onde  $S$  é um conjunto “de sentenças abstratas” e  $\vdash: P(S) \rightarrow S$  é um operador de consequência abstrato: este é o ponto de vista desenvolvido em [24], que se concentra no estudo das traduções conservativas abstratas. Alternativamente esta informação pode ser codificada por um par  $(S, \mathcal{T})$  dado por conjunto (de “sentenças”) munido de um operador de fecho definido em seu conjunto de partes  $\mathcal{T}: P(S) \rightarrow P(S)$ : este é o ponto de partida de [41], que apresenta uma variação interessante que incorpora a propriedade de estruturalidade de uma relação de consequência abstrata, trazendo a campo a noção matemática de módulo sobre um quantale para produzir novas noções de identidade de lógicas.

Por um breve momento, vamos utilizar uma situação específica: a da Lógica Clássica *de Predicados de primeira ordem* - como motivação para algumas abstrações. Aqui cabe reconhecermos que a teoria de modelos desta lógica “funciona bem” devido, em grande parte, a dois aspectos: (i) por um lado, em função de (meta)teoremas estruturais específicos



como o Teorema da Compacidade e o Teorema de Löwenheim-Skolem; (ii) por outro lado, devido a motivos gerais como as construções envolvendo restrições e expansões de linguagem com adições de novos símbolos para constantes e símbolos funcionais. Ambos aspectos admitem generalizações interessantes - a Teoria das Lógicas Abstratas e a Teoria das Instituições- constituindo instâncias de um novo campo de estudos denominado Teoria Abstrata de Modelos, que vamos apresentar brevemente na sequência.

Uma lógica abstrata é uma família de pares ordenados  $L = (S_\tau, \models_\tau)_\tau$  indexada pelas linguagens ou vocabulários  $\tau$ , onde  $S_\tau$  é um conjunto (“de  $\tau$ -sentenças abstratas”) e  $\models_\tau$  é uma relação (“de consequência semântica”) entre  $\tau$ -estruturas e  $\tau$ -sentenças, satisfazendo algumas condições de compatibilidade. O primeiro Teorema de Lindström caracteriza a Lógica de primeira ordem clássica como a lógica abstrata *regular* - representando, em particular, os conectivos booleanos - que é maximal entre as que satisfazem a propriedade de compacidade e de Löwenheim-Skolem para baixo: para uma prova deste resultado veja por exemplo o Capítulo XIII de [21]; para variantes deste resultado veja por exemplo [23].

A noção de *Instituição* foi introduzida por J. Goguen e R. Burstall no final da década de 1970 a fim de providenciar uma unificação formal do conceito de sistema lógico, motivados pela explosão populacional de lógicas e de suas formas de apresentação. Assim este conceito formaliza a noção de sistema lógico como um objeto matemático, i.e., fornece uma teoria de modelo abstrata e categorial que formaliza a noção intuitiva de sistema lógico, incluindo sintaxe, semântica e relação de satisfação entre eles ([20]). A principal característica (modelo-teórica) é que uma instituição contém uma relação de satisfação entre modelos e sentenças (ambos abstratos) que são coerentes sob a mudança de linguagens ou vocabulários (abstratos) por meio de morfismos. Lógicas de predicados de primeira ordem (finitárias ou infinitárias), com a semântica de Tarski, fornecem exemplos naturais de instituições. Uma variação do formalismo das instituições, a noção de  $\Pi$ -*Instituição*, foi definida por J. Fiadeiro e A. Sernadas em [26], fornecendo uma abordagem (prova-teórica) alternativa a noção de instituição ao substituir uniformemente a noção de modelo e satisfação por um operador de consequência primitivo à lá Tarski. Em [37] é verificado que as categorias naturais formadas por lógicas proposicionais,  $\mathcal{L}_s$  e  $\mathcal{L}_f$ , fornecem exemplos naturais de  $\Pi$ -instituições. Neste mesmo trabalho, mostra-se que, com noções naturais de flechas entre instituições e entre  $\Pi$ -instituições, obtem-se categorias (grandes),  $INST$  e  $\Pi - INST$ , e também um par de funtores adjuntos entre estas categorias. Já em [35], é fornecida uma *instituição* para cada categoria de lógicas proposicionais,  $\mathcal{L}_s$  e  $\mathcal{L}_f$ , através do uso da noção de matriz para uma lógica proposicional. Neste mesmo trabalho, por uma conveniente modificação da construção anterior, determina-se uma instituição para cada classe de equipolência de lógicas algebrizáveis: isso permite aplicar noções e resultados da Teoria da Instituições em Lógica Proposicional para derivar, a partir da introdução da noção de “contexto de Glivenko”, uma forma forte e geral do Teorema de Glivenko, que relaciona um par bem comportado de lógicas proposicionais (como o formado pela lógica clássica e a lógica intuicionista) obtendo assim um par de traduções conservativas entre as lógicas em questão a partir de um morfismo institucional entre as instituições lógicas correspondentes.

Este último resultado indica a utilidade potencial de não apenas considerar versões abstratas de lógicas (sejam proposicionais ou não) mas também considerar formas menos diretas de representar lógicas. Nesta linha aparecem os trabalhos [36], [4] e [34], que descreveremos brevemente na sequência.

Em [36] são estabelecidos códigos funtoriais perfeitos tanto para morfismos flexíveis de assinaturas quanto para morfismos flexíveis entre lógicas algebrizáveis. Em mais detalhes: (i) é estabelecido um anti-isomorfismo entre a classe de morfismos de assinaturas e alguns funtores entre as respectivas categorias de estruturas; (ii) prova-se que este anti-isomorfismo se restringe a um anti-isomorfismo entre morfismos de lógicas algebrizáveis e alguns funtores entre suas categorias de estruturas que se restringem às suas quasivariiedades canônicas que são suas respectivas semânticas algébricas equivalentes.

A lacuna conceitual do trabalho [36]- a saber a ausência de um código funtorial perfeito para lógicas proposicionais gerais - é preenchida pelo trabalho [4]. Isto é realizado através da noção de par filtro, um conceito introduzido em [39], como uma ferramenta para criar e analisar lógicas. Em mais detalhes: um par filtro para uma assinatura  $\Sigma$  é um par  $(G, j)$  dado por funtor contravariante  $G$  da categoria das  $\Sigma$ -estruturas na categoria dos reticulados algébricos e por uma transformação natural  $j : G \rightarrow P$ , onde  $P : \Sigma\text{-Str} \rightarrow AL$  é o funtor contravariante dado pelo conjunto potência do conjunto subjacente a uma estrutura, e tal que as componentes de  $j$  são funções que preservam ínfimos arbitrários e supremos dirigidos. É mostrado que toda lógica tarskiana surge de um par filtro e que as traduções flexíveis de lógicas surgem de morfismos de pares filtros, estabelecendo uma conexão categorial precisa: a categoria  $\mathcal{L}_f$  é isomorfa a uma subcategoria plena e coreflexiva da categoria formada por pares filtros e morfismos adequados de “mudança de assinatura base”.

Os trabalhos mencionados nos dois parágrafos anteriores apresentam um traço comum: produzir códigos (funtoriais) *perfeitos* de lógicas. Em Matemática, existem várias instâncias úteis de codificações mais relaxadas, como por exemplo quando se representam grupos por grupos de matrizes ou quando se identificam anéis com unidade  $R, R'$  que são Morita-equivalentes, i.e., quando as respectivas categorias de representações lineares (categorias de módulos) são equivalentes:  $R\text{-Mod} \simeq R'\text{-Mod}$ . Esta última relação está longe de constituir uma representação fiel da identidade estrutural de anéis<sup>9</sup> e daí segue sua utilidade. Em [34] são dados os primeiros passos para se constituir uma teoria de representações de lógicas: (i) seleciona-se uma classe de objetos padrões “bem comportados”, a classe das lógicas (Lindenbaum) algebrizáveis, que têm uma boa representação categorial (dada pelas quasivariiedades canonicamente associada a essas); (ii) por meio desta classe de objetos especiais define-se o conceito de lógicas *Morita equivalentes* (à esquerda) e uma variante mais fraca denominada lógicas estavelmente Morita equivalentes (à esquerda); (iii) mostra-se que uma lógica e sua lógica congruencial canonicamente associada são Morita equivalentes (apesar de, em geral, não serem equipolentes); (iv) mostra-se que as diversas apresentações da lógica clássica são Morita equivalentes; (v) verifica-se que a lógica clássica e a lógica intuicionista *não são* estavelmente Morita equivalentes.

Finalizamos estas notas com algumas palavras sobre identidades de ordem alta entre lógicas. Anteriormente nos focamos nas apresentações (e representações) de sistemas lógicos sob um ponto de vista prova-teórico, e também o mais básico entre esses: baseadas essencialmente na apresentação (ou codificação) de apenas uma relação de demonstrabilidade, eventualmente abstrata. Na realidade esse está longe de ser o espírito principal da subárea da Lógica denominada “Teoria da Prova”, que como o próprio nome coloca, foca

<sup>9</sup> I.e., existem anéis Morita equivalentes que não são isomorfos. Por exemplo, os anéis  $R$  e  $Mat_{n \times n}(R)$  são sempre Morita equivalentes, para qualquer  $n \geq 1$ , mas mesmo que  $R$  seja comutativo, o anel  $Mat_{n \times n}(R)$  nunca é comutativo se  $n \geq 2$ , logo estes não podem ser anéis isomorfos.

seu estudo nas provas em si e não na relação de demonstrabilidade (i.e. se existe uma prova de uma fórmula (eventualmente abstrata) a partir de algum conjunto de fórmulas). Assim, por esse ponto de vista, seria muito natural considerar identidades entre provas como parte constituinte de alguma conceituação de identificação de lógicas: por exemplo, colocando que duas provas são indistinguíveis quando se reduzem à mesma forma normal. A situação seria, grosso modo, assim descrita: a cada lógica atribuímos uma multi-categoria ([30]) cujos objetos são as fórmulas e cujas multi-flechas, que tem fonte um conjunto de fórmulas e alvo uma fórmula única, são as provas desta fórmula com hipóteses no conjunto; a composição de multi-flechas seria dada pela concatenação de provas (como esperado); as multi-identidades seriam dadas pela propriedade de reflexividade<sup>10</sup>. Assim os sistemas lógicos anteriormente considerados seriam simplesmente “multi-posets”, uma versão “magra” da multi-categoria acima descrita: i.e. haveria uma *única* multi-flecha de um conjunto de fórmulas em uma fórmula sempre que existisse alguma prova da fórmula com hipóteses neste conjunto. Assim, sob o ponto de vista anteriormente discutido, lógicas seriam indistinguíveis quando os multi-posets associados fossem equivalentes. Por outro lado, na multi-categoria cheia, cujas multi-flechas são as provas, haveria muito mais estrutura a ser considerada, tendo assim noções muito mais finas de indistinguibilidade de lógicas: lógicas seriam indistinguíveis se, além de haver um processo de identificação dos objetos dados por (classes de equivalência) de fórmulas, este processo também respeitasse (e identificasse) a estrutura local de provas entre as fórmulas associadas.

Tendo em vista as colocações acima e trazendo também a campo a essência da correspondência de Curry-Howard, que conecta lógicas com teorias de tipos - identificando fórmulas com tipos e assim também identificando provas de fórmulas com termos de um dado tipo - fica natural entendermos as iniciativas que surgiram a partir de meados da década de 1990 de conectar teoria de tipos e teoria da homotopia abstrata ([29]). Assim a questão de identidades entre provas ficaria codificada através da família, indexada nos tipos, formada pelos tipos de identidade associado a cada tipo, que codifica a estrutura das noções de indistinguibilidade entre termos de um dado tipo. Assim as próprias identidades entre termos (provas) de um dado tipo (fórmula) seriam dados por “homotopias formais” entre esses termos. Por outro lado, estas homotopias formais poderiam ser realizados concretamente por caminhos computacionais (reversíveis) entre provas (= multi-flechas) dados por sequências finitas de substituições e reescrituras que transformariam uma prova na outra prova. E também poderíamos imaginar identidades de ordem ainda mais alta, por exemplo, observando a estrutura das homotopias formais entre homotopias formais, e assim por diante.

No espírito das observações colocadas nos parágrafos acima, finalizamos essas notas simplesmente sugerimos ao leitor interessado a consulta das referências [29], [28], [2] e [40].

<sup>10</sup> Toda fórmula determina uma prova dela mesma.

## Referências Bibliográficas

- [1] P. Arndt, R. A. Freire, O. O. Luciano, H. L. Mariano, *A global glance on categories in Logic*, *Logica Universalis* **1** (2007), 3-39.
- [2] P. Arndt, C. Kapulkin, *Homotopy Theoretic Models of Type Theory*, arXiv preprint, <https://arxiv.org/abs/1208.5683>, 2012.
- [3] E. L. B. de Almeida, **Análise das condições de verdade e dos requerimentos existenciais em axiomatizações da aritmética**, Tese de Doutorado em Filosofia, IFCH-Unicamp, 2017.
- [4] P. Arndt, H. L. Mariano, D. C. Pinto, *Finitary Filter Pairs and Propositional Logics*, a aparecer em *South American Journal of Logic*, 2019.
- [5] P. Arndt, *Homotopical Categories of Logics*, em **The Road to Universal Logic** (2015), 13-58.
- [6] J. Adámek, J. Rosický, **Locally Presentable and Accessible Categories**, Lecture Notes Series **189**, London Math. Soc., Cambridge University Press, 1994.
- [7] J. Bueno-Soler, W. A. Carnielli, *Possible-translations algebraization for paraconsistent logics*, *Bulletin of the Section of Logic*, University of Lodz, Poland, vol. 34, n. 2, 2005, pp. 77-92. *CLE e-Prints vol.5 n.6*, (2005), <http://www.cle.unicamp.br/e-prints>.
- [8] J. Bueno, M. E. Coniglio, W. A. Carnielli, *Finite algebraizability via possible-translations semantics*, *Proceedings of CombLog'04 — Workshop on Combination of Logics: Theory and Applications*, Departamento de Matemática, Instituto Superior Técnico, Lisboa, Portugal, 2004.
- [9] J. Bueno-Soler, M. E. Coniglio, W. A. Carnielli, *Possible-Translations Algebraizability*, **Paraconsistency with no Frontiers** (J.-Y. Béziau, W. Carnielli, eds.), North-Holland, 2006.
- [10] J.-Y. Béziau, *From Consequence Operator to Universal Logic: A Survey of General Abstract Logic*, em **Logica Universalis: Towards a General Theory of Logic** (J.-Y. Béziau, ed.), Birkhäuser, 2007, 3-17.
- [11] J.-Y. Béziau, R. P. de Freitas, J. P. Viana, *What is Classical Propositional Logic? (A Study in Universal Logic)*, *Logica Studies* **7**, 2001.
- [12] W. J. Blok, D. Pigozzi, **Algebraizable logics**, *Memoirs of the AMS* **396**, American Mathematical Society, Providence, USA, 1989.
- [13] W. A. Carnielli, *Many-valued logics and plausible reasoning*, *Proceedings of the XX International Congress on Many-Valued Logics*, IEEE Computer Society, University of Charlotte, USA, 1990, 328-335.
- [14] W. A. Carnielli, M. E. Coniglio, *Combining Logics*, *Stanford Encyclopedia of Philosophy*, <http://plato.stanford.edu/entries/logic-combining/>, 2016.
- [15] C. Caleiro, W. Carnielli, J. Rasga, C. Sernadas, *Fibring of Logics as a Universal Construction*, *Handbook of Philosophical Logic* **13** (editores: D. Gabbay, F. Guentner), Kluwer Academic Publishers, 2005.
- [16] C. Caleiro, R. Gonçalves, *Equipollent logical systems*, em **Logica Universalis: Towards a General Theory of Logic** (J.-Y. Béziau ed.), 2007, 97-110.
- [17] M. E. Coniglio, *The Meta-Fibring environment: Preservation of meta-properties by fibring*, *CLE e-Prints* **5(4)** (2005).
- [18] C. Caleiro, J. Ramos, *Cryptofibring*, *Proceedings of CombLog 04 - Workshop on Combination of Logics: Theory and Applications*, Lisboa, Portugal (editores: W. A. Carnielli, F. M. Dionísio, P. Mateus), 2004, 87-92.
- [19] J. Czelakowski, **Protoalgebraic logic**, *Trends in Logic*, *Studia Logica Library*, Kluwer Academic Publishers, 2001.
- [20] R. Diaconescu, **Institution-independent Model Theory**, Birkhäuser, 2000.
- [21] H. D. Ebbinghaus, J. Flum, W. Thomas, **Mathematical Logic**, *Undergraduate Texts in Mathematics*, segunda edição, Springer-Verlag, 1994.

- 
- [22] V. L. Fernández, M. E. Coniglio, *Fibring algebraizable consequence systems*, Proceedings of CombLog 04 - Workshop on Combination of Logics: Theory and Applications, (editores: W. A. Carnielli, F. M. Dionísio, P. Mateus), 2004, 93-98.
- [23] J. Flum, *Characterizing Logics*, Capítulo 2 em **Model-Theoretic Logics** (ed. J. Barwise, S. Feferman), Perspectives in Mathematical Logic, Springer-Verlag, 1985.
- [24] H. A. Feitosa, I. M. L. D'Ottaviano, *Conservative translations*, Annals of Pure and Applied Logic **108** (2001), 205-227.
- [25] A. R. de O. Freire Filho, **Estudo Comparado do Comprometimento Ontológico das Teorias de Classes e Conjuntos**, Tese de Doutorado em Filosofia, IFCH-Unicamp, 2019.
- [26] J. Fiadeiro, A. Sernadas, *Structuring theories on consequence*, Recent Trends in Data Type Specification (D. Sannella, A. Tarlecki eds.), Lecture Notes in Comput. Sci. **332** (1988), 44-72.
- [27] D. Gabbay, *Fibred semantics and the weaving of logics: Part 1*, Journal of Symbolic Logic **61**(4) (1996), 1057-1120.
- [28] **Homotopy Type Theory: Univalent Foundations of Mathematics**, Institute for Advanced Study, disponível em <https://homotopytypetheory.org/book/>.
- [29] M. Hofmann, T. Streicher, *The groupoid interpretation of type Theory*, em **Twenty-Five Years of Constructive Type Theory (Venice, 1995)**, Oxford Logic Guides **36**, Oxford University Press, 1998, 83-111.
- [30] T. Leinster, **Higher operads, higher categories**, Lecture Notes Series **298**, London Math. Soc., Cambridge University Press, 2004.
- [31] J. Łoś, R. Suszko, *Remarks on sentential logics*, Proceedings Koninklijke Nederlandse Akademie van Wetenschappen, Series A, **61** (1958), 177-183.
- [32] S. MacLane, **Categories for the Working Mathematician**, Graduate Texts in Mathematics **5**, segunda edição, Springer-Verlag, 1998.
- [33] H. L. Mariano, C. A. Mendes, *Towards a good notion of categories of logics*, arXiv preprint, <http://arxiv.org/abs/1404.3780>, 2016.
- [34] H. L. Mariano, D. C. Pinto, *Representation theory of logics: a categorical approach*, arXiv preprint, <http://arxiv.org/abs/1405.2429>, 2014.
- [35] H. L. Mariano, D. C. Pinto, *An abstract approach to Glivenko's theorem*, arXiv preprint, <http://arxiv.org/abs/1612.03410>, 2016.
- [36] H. L. Mariano, D. C. Pinto, *Algebraizable Logics and a functorial encoding of its morphisms*, Logic Journal of the IGPL **25**(4) (2017), 524-561.
- [37] H. L. Mariano, D. C. Pinto, *Remarks on Propositional Logics and the Categorical Relationship Between Institutions and  $\Pi$ -Institutionst*, South American Journal of Logic **3**(1) (2017), 111-121.
- [38] H. L. Mariano, **Grupos Especiais, Teoria dos Modelos e Categorias de Lógicas**, Tese de Livre-Docência, IME-USP, 2011.
- [39] D. C. Pinto, **A categorical foundation for a representation theory of logics**, Tese de Doutorado em Matemática, IME-USP, 2016.
- [40] R. J. G. B. de Queiroz, A. G. de Oliveira, A. F. Ramos, *Propositional Equality, Identity Types, and Computational Paths*, South American Journal of Logic **2**(2) (2016), 245-296.
- [41] C. Russo, *An order-theoretic analysis of interpretations among propositional deductive systems*, Annals of Pure and Applied Logic **164** (2013), 112-130.
- [42] C. Sernadas, J. Rasga, W. A. Carnielli, *Modulated fibring and the collapsing problem*, The Journal of Symbolic Logic **67** (2002), 1541-1569.
- [43] A. Sernadas, C. Sernadas, C. Caleiro, *Fibring of logics as a categorical construction*, Journal of Logic and Computation **9**(2) (1999), 149-179.

# CATEGORIAS COM MORFISMOS VERDADE E PROTOTOPOS: UMA SEMÂNTICA PARA LINGUAGENS PROPOSICIONAIS<sup>1</sup>

Edelcio G. de Souza<sup>2</sup>

## Abstract

O objetivo do presente trabalho é desenvolver algumas propriedades de categorias que são mais gerais que categorias que são *topos*. Apresentamos, assim, os conceitos de *categorias com morfismos verdade* (CTM) e *prototopos*, mostrando como se podem definir os morfismos verdade nessas categorias de modo a prover uma semântica para linguagens proposicionais usuais. Mostramos, também, que em categorias com morfismos verdade vale um teorema análogo ao teorema fundamental da teoria de topos.

## 1. Introdução

A fim de acompanhar o presente trabalho, assumimos que o leitor esteja familiarizado com os conceitos básicos de teoria de categorias, incluindo a noção de limite e colimite, classificadores de subobjetos e exponenciação<sup>3</sup>. Esperamos, assim, que o leitor conheça as propriedades fundamentais de categorias que são *topos*<sup>4</sup>. O objetivo é introduzir um tipo mais geral de categoria, que denominaremos *prototopos*, que constitua um ambiente em que se possa construir semântica para linguagens proposicionais<sup>5</sup>. Precisamos, então, de um tipo de categoria em que os valores de verdade podem ser definidos junto com os morfismos verdade que correspondem às funções de verdade usuais como negação, conjunção, implicação e disjunção. Vamos mostrar que para estas construções não precisamos que a categoria seja cartesiana fechada, em particular que tenha exponenciação, mas apenas que ela satisfaça uma propriedade dessas categorias, a saber, o que denominamos de **0-Axioma**: *se um morfismo possui como codomínio um objeto inicial, então ele é um isomorfismo*.

Segue-se do 0-Axioma, que se o domínio de um morfismo é um objeto inicial, então ele deve ser um monomorfismo. Isso implica que o morfismo

$$0 \rightarrow 1$$

deve ser um monomorfismo. Assim, assumindo que a categoria possui classificador de subobjetos  $\top : 1 \rightarrow \Omega$ , o morfismo acima possui um caráter dado por  $\perp : 1 \rightarrow \Omega$ , que é identificado como o valor de verdade *falso*.

Se a categoria é não degenerada, temos assim dois valores de verdade diferentes,  $\top, \perp : 1 \rightarrow \Omega$ , que correspondem ao verdadeiro e ao falso.

---

<sup>1</sup> Esse trabalho foi desenvolvido durante um estágio de pesquisa na Universidade de Miami. Agradeço ao Prof. Dr. Otávio Bueno por sua hospitalidade e à Fundação de Amparo à Pesquisa do Estado de São Paulo pela concessão da bolsa de pesquisa no exterior, processo número FAPESP 2018/10326-4.

<sup>2</sup> Departamento de Filosofia, FFLCH, USP. Email: [edelcio.souza@usp.br](mailto:edelcio.souza@usp.br).

<sup>3</sup> Vamos utilizar a notação de Goldblatt [6].

<sup>4</sup> O leitor pode consultar além de Goldblatt [6], os seguintes textos fundamentais: Awodey [1], Bell [2], Herrlich & Strecker [7] e MacLane [9]. Para a teoria de topos ver: MacLane & Moerdijk [10] e MacLarty [11].

<sup>5</sup> Uma versão preliminar e resumida do presente trabalho pode ser vista em de Souza [4].

Com o falso, é fácil construir a função de verdade que corresponde à negação. Como  $\perp : 1 \rightarrow \Omega$  é um monomorfismo, seu caráter é dado por  $\chi_{\perp} = f_{\neg} : \Omega \rightarrow \Omega$  que é o morfismo verdade que corresponde à negação. Para ver isso, é suficiente mostrar que  $f_{\neg} \circ \top = \perp$  e  $f_{\neg} \circ \perp = \top$  e temos o comportamento usual da função de verdade que corresponde à negação com respeito aos valores de verdade verdadeiro e falso.

Vamos mostrar que podemos fazer as mesmas construções para os morfismos que correspondem às outras funções usuais de verdade, e apresentar como se comportam com respeito aos valores de verdade. Para isto definiremos, por meio de axiomas, dois tipos de categorias em que todas essas construções podem ser realizadas.

## 2. Categorias com morfismos verdade

Dizemos que uma categoria  $\mathcal{E}$  é uma *categoria com morfismos verdade* (CTM) se e somente se ela satisfaz os seguintes axiomas:

(CTM1)  $\mathcal{E}$  é finitamente completa;

(CTM2)  $\mathcal{E}$  é finitamente cocompleta;

(CTM3)  $\mathcal{E}$  possui classificador de subobjetos;

(CTM4)  $\mathcal{E}$  satisfaz a seguinte propriedade: (0-Axioma) Se existe um  $\mathcal{E}$ -morfismo  $f : a \rightarrow 0$  ( $0$  é inicial em  $\mathcal{E}$ ),  $f$  é um isomorfismo. E, então, temos que  $a$  e  $0$  são isomorfos, e  $a$  é também inicial em  $\mathcal{E}$ .

Uma categoria  $\mathcal{E}$  é dita *finitamente bicompleta* (completa e cocompleta) quando todo diagrama finito em  $\mathcal{E}$  possui limite e colimite. Nesse caso,  $\mathcal{E}$  possui objetos terminais e iniciais e, em  $\mathcal{E}$ , pode-se fazer uma série de construções úteis: produtos e coprodutos de objetos e morfismos, equalizadores e coequalizadores, pullbacks e pushouts, etc.

Considere um objeto  $d$  em uma categoria  $\mathcal{E}$ . Um *subobjeto* de  $d$  é uma certa classe de equivalência de monomorfismos com codomínio  $d$ . Sejam os  $\mathcal{E}$ -monomorfismos  $f : a \rightarrow d$  e  $g : b \rightarrow d$ . Dizemos que  $f$  é *equivalente* a  $g$ , em símbolos  $f \equiv g$  se existe um isomorfismo  $h : a \rightarrow b$  tal que o diagrama

$$\begin{array}{ccc} a & \xrightarrow{f} & d \\ h \downarrow & & \nearrow g \\ b & & \end{array}$$

comuta, isto é,  $f = g \circ h$ .

É fácil ver que  $\equiv$  é uma relação de equivalência e, se  $Mon(d)$  é a classe dos monomorfismos com domínio  $d$ , então um *subobjeto* de  $d$  é qualquer elemento do conjunto quociente:

$$Mon(d)/_{\equiv} := \{[f]_{\equiv} : f \in Mon(d)\}.$$

Se  $\mathcal{E}$  é uma categoria com objeto terminal  $1$ , então um *classificador de subobjetos* para  $\mathcal{E}$  é um  $\mathcal{E}$ -objeto  $\Omega$  junto com um  $\mathcal{E}$ -morfismo

$$\top : 1 \rightarrow \Omega$$

tal que satisfaz a seguinte propriedade universal:

( $\Omega$ -Axioma): Para cada monomorfismo  $f : a \rightarrow b$  existe um único  $\mathcal{E}$ -morfismo  $\chi_f : d \rightarrow \Omega$ , denominado o *morfismo característico* ou o *caráter* de  $f$ , tal que o diagrama

$$\begin{array}{ccc} a & \xrightarrow{f} & d \\ \downarrow & & \downarrow \chi_f \\ 1 & \xrightarrow{\top} & \Omega \end{array}$$

é um quadrado pullback.

Uma categoria  $\mathcal{E}$  é dita um *topos* se e somente se  $\mathcal{E}$  é finitamente bicompleta, possui classificador de subobjetos e exponenciação.

Diz-se que uma categoria  $\mathcal{E}$  possui exponenciação se, para cada  $\mathcal{E}$ -objeto  $a$ , temos um funtor exponencial  $(-)^a : \mathcal{E} \rightarrow \mathcal{E}$  que é adjunto à direita do funtor  $(-) \times a$ .

Em termos elementares, isso significa que  $\mathcal{E}$  tem produto para cada par de  $\mathcal{E}$ -objetos e, para quaisquer  $\mathcal{E}$ -objetos dados  $a$  e  $b$ , existe um  $\mathcal{E}$ -objeto, dito a *exponencial* de  $b$  por  $a$ , denotado por  $b^a$ , e um  $\mathcal{E}$ -morfismo  $ev_{ab} : b^a \times a \rightarrow b$ , denominado *morfismo avaliação*, tal que vale a seguinte propriedade universal: Para todo  $\mathcal{E}$ -objeto  $c$  e  $\mathcal{E}$ -morfismo  $g : c \times a \rightarrow b$ , existe um único  $\mathcal{E}$ -morfismo  $\hat{g} : c \rightarrow b^a$  tal que o seguinte diagrama comuta:

$$\begin{array}{ccc} & b^a \times a & \\ & \uparrow & \searrow ev_{ab} \\ \hat{g} \times 1_a & & b \\ & \downarrow & \nearrow g \\ & c \times a & \end{array}$$

isto é,  $ev_{ab} \circ (\hat{g} \times 1_a) = g$ . Segue-se, imediatamente, que existe uma correspondência biunívoca entre as classes  $\mathcal{E}(c \times a, b)$  e  $\mathcal{E}(c, b^a)$ . (Se  $\mathcal{E}$  é uma categoria e  $a$  e  $b$  são dois  $\mathcal{E}$ -objetos, então  $\mathcal{E}(a, b)$  denota a classe de todos os  $\mathcal{E}$ -morfismos de  $a$  em  $b$ .) Quando uma categoria  $\mathcal{E}$  é finitamente completa e possui exponenciação, então  $\mathcal{E}$  é denominada uma *categoria cartesiana fechada*.

Dado que as categorias cartesianas fechadas possuem como consequência o 0-Axioma, segue-se que *todo topos é uma categoria com morfismos verdade* mas, no entanto, a converso é falsa. Um exemplo de uma CTM que não é um topos é a categoria  $SET^{\leq \aleph_0}$ , dos conjuntos no máximo enumeráveis e funções entre esses conjuntos, que satisfaz o 0-Axioma mas não possui exponenciação.

### 3. Construções básicas em CTM's

Vamos começar a examinar que construções e propriedades que são válidas em topos e que permanecem válidas nas categorias com morfismos verdade.

Seja  $\mathcal{E}$ , então, uma CTM.

PROPOSIÇÃO 1. *Todo  $\mathcal{E}$ -morfismo do tipo  $0 \xrightarrow{f} a$  é um monomorfismo.*

**Prova.** De fato, suponha que existam morfismos  $g, h : x \rightarrow 0$  tais que  $f \circ g = f \circ h$ . Pelo 0-Axioma,  $g$  e  $h$  são isomorfismos e, assim,  $x$  é inicial em  $\mathcal{E}$ . Portanto,  $g = h$  e  $f$  é monomorfismo.  $\square$

PROPOSIÇÃO 2. *Todo monomorfismo em  $\mathcal{E}$  é um equalizador.*

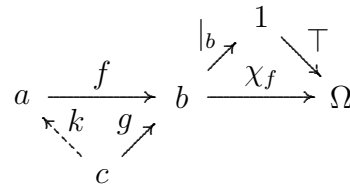
**Prova.** Seja  $f : a \rightarrow b$  um monomorfismo em  $\mathcal{E}$ . Pelo  $\Omega$ -Axioma, temos o seguinte pullback:

$$\begin{array}{ccc} a & \xrightarrow{f} & b \\ |a \downarrow & \swarrow |b & \downarrow \chi_f \\ 1 & \xrightarrow{\top} & \Omega \end{array}$$

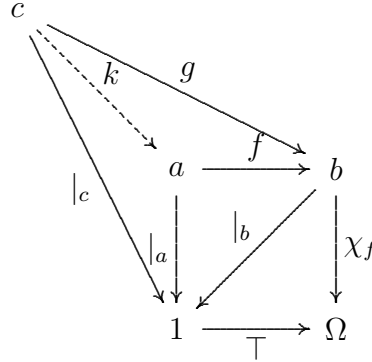
Queremos mostrar que  $f$  é um equalizador de  $\chi_f$  e  $\top_b := \top \circ |b$ . Verifiquemos primeiramente que  $\chi_f \circ f = \top_b \circ f$ . Como  $1$  é terminal em  $\mathcal{E}$ , temos que  $|a = |b \circ f$  e, assim,



$$\chi_f \circ f = \top \circ |_a = \top \circ |_b \circ f = \top_b \circ f.$$



Seja, então,  $g : c \rightarrow b$  tal que  $\chi_f \circ g = \top_b \circ g$ . (Ver diagrama acima.)



Como 1 é terminal,  $|_c = |_b \circ g$ . Assim,  $\chi_f \circ g = \top_b \circ g = \top \circ |_b \circ g = \top \circ |_c$  e, portanto, o perímetro do diagrama acima comuta. Como se trata de um pullback, existe um único morfismo  $k : c \rightarrow a$  tal que  $f \circ k = g$ . Assim,  $f$  é um equalizador de  $\chi_f$  e  $\top_b$ .  $\square$

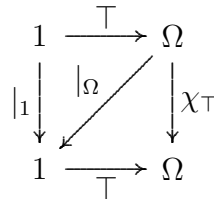
A proposição 2 possui dois corolários imediatos.

**COROLÁRIO 3.** *Em  $\mathcal{E}$ , um morfismo é isomorfismo se e somente se ele é monomorfismo e epimorfismo.*

**Prova.** De fato, vale em toda categoria que todo isomorfismo é monomorfismo e epimorfismo. Por outro lado, em  $\mathcal{E}$ , um morfismo que é epimorfismo e monomorfismo, pela proposição 2, é um epimorfismo equalizador. Mas, em uma categoria qualquer, um epimorfismo equalizador é sempre um isomorfismo.  $\square$

**COROLÁRIO 4.** *Em  $\mathcal{E}$ , um classificador de subobjetos  $\top : 1 \rightarrow \Omega$  é um equalizador de  $1_\Omega$  e  $\top_\Omega$ .*

**Prova.** Considere o seguinte diagrama:



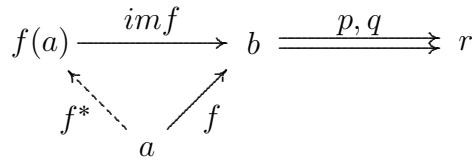
e o resultado é consequência imediata da proposição 2, lembrando que  $\chi_\top = 1_\Omega$ .  $\square$

Vamos, agora, considerar a importante construção de imagens de morfismos com a correspondente demonstração que em CTM's todo morfismo possui uma decomposição epi-mono. É importante perceber que em nenhuma destas construções é necessário que a categoria possua exponenciação.

Considere a categoria com morfismos verdade  $\mathcal{E}$  e seja  $f : a \rightarrow b$  um  $\mathcal{E}$ -morfismo qualquer. Como  $\mathcal{E}$  é finitamente cocompleta, forme o pushout:

$$\begin{array}{ccc} a & \xrightarrow{f} & b \\ f \downarrow & & \downarrow q \\ b & \xrightarrow{p} & r \end{array}$$

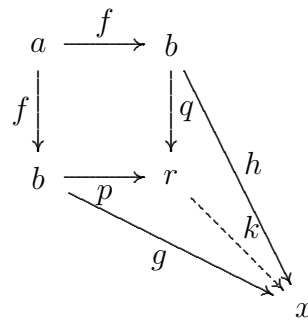
e seja  $imf : f(a) \twoheadrightarrow b$  o equalizador<sup>6</sup> de  $p$  e  $q$ . O morfismo  $imf$  é denominado o *morfismo imagem* de  $f$ . Como, pelo diagrama do pushout,  $q \circ f = p \circ f$  e  $imf$  é um equalizador, existe um único  $f^*$  tal que  $f = imf \circ f^*$ . (Ver diagrama abaixo.)



Vamos mostrar que  $f^*$  é um epimorfismo e como equalizadores são monomorfismos,  $f$  possui uma fatorização epi-mono. Para isto, vamos considerar primeiramente um resultado preliminar.

LEMA 5. *No pushout acima de  $f$  por  $f$ , se  $p = q$ , então  $f$  é epimorfismo.*

**Prova.** Considere os morfismos  $g, h : b \rightarrow x$  tais que  $g \circ f = h \circ f$ .

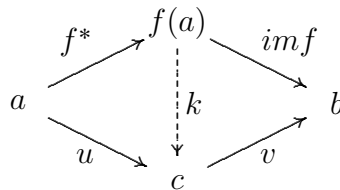


Como o quadrado é um pushout, existe um único  $k$  tal que  $k \circ p = g$  e  $k \circ q = h$ . Como  $p = q$ , temos:  $g = k \circ p = k \circ q = h$  e, assim,  $f$  é epimorfismo.  $\square$

PROPOSIÇÃO 6. *Seja  $f : a \rightarrow b$  um  $\mathcal{E}$ -morfismo. Assim,  $imf$  é o menor subobjeto de  $b$  para o qual  $f$  se fatora. Isto é, se há uma outra fatoração de  $f$  dada por*

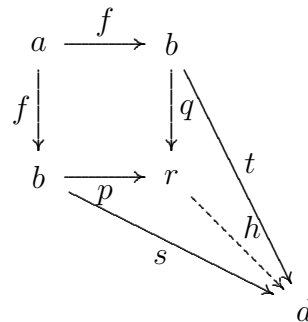
$$f = v \circ u : a \rightarrow c \twoheadrightarrow b$$

para algum  $u$  e  $v$  monomorfismo, então existe um único  $k : f(a) \rightarrow c$  que faz o diagrama a seguir comutar



e, assim,  $imf$  está incluída em  $v$ .

**Prova.** Como  $v$  é um monomorfismo, pela proposição 2,  $v$  é um equalizador de um par de morfismos  $s, t : b \rightarrow d$ . Considere o diagrama do pushout de  $f$  por  $f$ :



<sup>6</sup> Como limites e colimites de diagramas são sempre isomorfos, vamos usar o artigo definido fixando um equalizador específico qualquer. O mesmo abuso de linguagem será feito para todos os outros limites e colimites considerados no texto.

Então,  $s \circ f = s \circ v \circ u = t \circ v \circ u = t \circ f$ . Assim, existe um único  $h$  tal que  $h \circ p = s$  e  $h \circ q = t$ .

Portanto, temos que:

$$s \circ imf = h \circ p \circ imf = h \circ q \circ imf = t \circ imf.$$

Como  $v$  equaliza  $s$  e  $t$ , temos:

$$\begin{array}{ccc} c & \xrightarrow{v} & b \xrightarrow{s, t} r \\ & \swarrow k & \nearrow imf \\ & & f(a) \end{array}$$

Logo, existe um único  $k : f(a) \rightarrow c$  tal que  $v \circ k = imf$ . Resta mostrar que  $k \circ f^* = u$ . Mas,  $v \circ k \circ f^* = imf \circ f^* = f = v \circ u$ . Assim,  $v \circ k \circ f^* = v \circ u$ . Como  $v$  é monomorfismo (cancelável à esquerda), segue-se que  $k \circ f^* = u$ , e temos o resultado.  $\square$

Temos três corolários da proposição 6.

**COROLÁRIO 7.**  $f^* : a \rightarrow f(a)$  é um epimorfismo.

**Prova.** Aplicando o processo de construção de imagens no próprio  $f^*$  temos:

$$\begin{array}{ccc} a & \xrightarrow{f^*} & f(a) \\ & \searrow f^{**} & \nearrow imf^* \\ & & f^*(a) \end{array}$$

Fazendo  $f^* = g$ , temos o seguinte diagrama comutativo:

$$\begin{array}{ccccc} & & g(a) & & \\ & g^* \nearrow & \downarrow img & \searrow imf \circ img & \\ a & & & & b \\ & f^* \searrow & \downarrow img & \nearrow imf & \\ & & f(a) & & \end{array}$$

Então,  $imf \circ img$  é monomorfismo (composição de monomorfismos). Como  $imf$  é cancelável à esquerda,  $img$  é o único morfismo que faz  $imf \circ img$  estar incluído em  $imf$ . Por outro lado, aplicando a proposição 6, temos que  $imf$  está incluído em  $imf \circ img$ . Assim,  $imf \cong imf \circ img$  e temos  $g(a) \cong f(a)$ .

Por definição,  $img$  é o equalizador de:

$$g(a) \xrightarrow{img} f(a) \xrightarrow{p, q} r$$

em que  $p$  e  $q$  são dados pelo pushout:

$$\begin{array}{ccc} a & \xrightarrow{g = f^*} & f(a) \\ \downarrow g = f^* & & \downarrow q \\ f(a) & \xrightarrow{p} & r \end{array}$$

Como  $p \circ img = q \circ img$  e  $img$  é epimorfismo (pois é iso), então  $p = q$ . Segue-se, então, do lema 5 acima, que  $f^*$  é um epimorfismo.  $\square$

COROLÁRIO 8.  $imf \circ f^* : a \rightarrow f(a) \rightarrow b$  é uma fatoração epi-mono de  $f$  que é única a menos de isomorfismos comutantes, isto é, se  $v \circ u : a \rightarrow c \rightarrow b$  é tal que  $f = v \circ u$ , então existe exatamente um morfismo  $k : f(a) \rightarrow c$  tal que o seguinte diagrama comuta:

$$\begin{array}{ccccc}
 & & f(a) & & \\
 & f^* \nearrow & & \searrow imf & \\
 a & & & & b \\
 & u \searrow & & \nearrow v & \\
 & & c & & 
 \end{array}$$

e  $k$  é um isomorfismo.

**Prova.** Segue-se da proposição 6 que  $k$  existe e é único. É preciso apenas mostrar que  $k$  é isomorfismo. Como  $v \circ k = imf$  é monomorfismo, segue-se que  $k$  é monomorfismo. Dualmente, como  $k \circ f^* = u$  é epimorfismo,  $k$  é epimorfismo também. Assim,  $k$  é epi e mono e, pela proposição 3,  $k$  é um isomorfismo.  $\square$

COROLÁRIO 9.  $f : a \rightarrow b$  é epimorfismo se e somente se existe  $g : f(a) \cong b$  tal que  $g \circ f^* = f$ .

**Prova.** Considere a fatoração de  $f : a \rightarrow b$  dada por:

$$imf \circ f^* : a \rightarrow f(a) \rightarrow b.$$

( $\Rightarrow$ ) Se  $f$  é epi, então  $imf$  é epi e como  $imf$  é mono, do corolário 3, temos que  $imf : f(a) \cong b$ .

( $\Leftarrow$ ) Se tivermos

$$g \circ f^* : a \rightarrow f(a) \cong b$$

com  $f = f^* \circ g$ , então como  $g$  é epi (pois é iso) e  $f^*$  é epi, segue-se que  $f$  é epimorfismo.  $\square$

#### 4. Valores de verdade em CTM's

A fim de construir semântica para linguagens proposicionais em categorias com morfismos verdade, precisamos primeiramente considerar como serão os valores de verdade neste tipo de categoria.

Se  $a$  é um  $\mathcal{E}$ -objeto e  $1$  é terminal em  $\mathcal{E}$ , um morfismo do tipo  $1 \rightarrow a$  é dito um *elemento* de  $a$ . Assim, considerando o classificador de subobjetos, morfismos do tipo  $1 \rightarrow \Omega$  são os elementos de  $\Omega$  denominados *valores de verdade*. O morfismo  $\top : 1 \rightarrow \Omega$  é dito o valor de verdade *verdadeiro*. Vejamos como obtemos o valor de verdade correspondente ao falso.

Pela proposição 1 acima, o morfismo  $0_1 : 0 \rightarrow 1$  é um monomorfismo. Aplicando o  $\Omega$ -Axioma, obtemos o caráter de  $0_1$  dado pelo pullback:

$$\begin{array}{ccc}
 0 & \xrightarrow{0_1} & 1 \\
 0_1 \downarrow & & \downarrow \chi_{0_1} \\
 1 & \xrightarrow{\top} & \Omega
 \end{array}$$

Definimos, então, o elemento de  $\Omega$  dado por  $\perp : 1 \rightarrow \Omega$ , como  $\perp := \chi_{0_1}$ . Denominamos o morfismo  $\perp$  o valor de verdade *falso*.

Assim, temos o pullback dado por:

$$\begin{array}{ccc}
 0 & \xrightarrow{0_1} & 1 \\
 0_1 \downarrow & & \downarrow \perp \\
 1 & \xrightarrow{\top} & \Omega
 \end{array}$$

Vejamos o que acontece se  $\top = \perp$ . Para isto, vamos considerar alguns fatos preliminares acerca de pullbacks que valem em categorias em geral.

FATO 10. Se o quadrado dado pelo diagrama

$$\begin{array}{ccc} a & \xrightarrow{p} & b \\ q \downarrow & & \downarrow f \\ b & \xrightarrow{f} & c \end{array}$$

é um pullback, então  $p$  e  $q$  são epimorfismos.

**Prova.** Considere o diagrama abaixo lembrando que identidades são isomorfismos (mono e epi).

$$\begin{array}{ccccc} & & & & b \\ & & & & \searrow 1_b \\ & & & & a \xrightarrow{p} b \\ & & & & \downarrow q \\ & & & & b \xrightarrow{f} c \\ & & & & \downarrow f \\ & & & & c \end{array}$$

*(Note: The diagram above is a simplified representation. The actual diagram in the image shows a pullback square with an additional object  $b$  at the top left. A dashed arrow  $h$  goes from  $b$  to  $a$ . Solid arrows  $1_b$  go from  $b$  to  $a$  and  $b$  to  $b$ . Solid arrows  $q$  and  $f$  go from  $a$  and  $b$  to  $b$  and  $c$  respectively. Solid arrows  $p$  and  $f$  go from  $a$  and  $b$  to  $b$  and  $c$  respectively.)*

Como a parte periférica do diagrama comuta, então existe um único  $h : b \rightarrow a$  tal que  $p \circ h = 1_b$  e  $q \circ h = 1_b$ . Como  $1_b$  é epi,  $p$  e  $q$  devem ser epimorfismos. □

FATO 11. Se o quadrado dado pelo diagrama

$$\begin{array}{ccc} a & \xrightarrow{p} & b \\ q \downarrow & & \downarrow f \\ b & \xrightarrow{f} & c \end{array}$$

é um pullback, então  $f$  é monomorfismo se e somente se  $p = q$ .

**Prova.** Suponha que  $f$  é monomorfismo. Como  $f \circ p = f \circ q$  e  $f$  é cancelável à esquerda, segue-se que  $p = q$ .

Suponha, agora, que  $p = q$  e considere o diagrama

$$\begin{array}{ccccc} & & & & d \\ & & & & \searrow h \\ & & & & a \xrightarrow{p} b \\ & & & & \downarrow q \\ & & & & b \xrightarrow{f} c \\ & & & & \downarrow f \\ & & & & c \end{array}$$

*(Note: The diagram above is a simplified representation. The actual diagram in the image shows a pullback square with an additional object  $d$  at the top left. A dashed arrow  $k$  goes from  $d$  to  $a$ . Solid arrows  $g$  and  $h$  go from  $d$  to  $a$  and  $b$  respectively. Solid arrows  $q$  and  $f$  go from  $a$  and  $b$  to  $b$  and  $c$  respectively. Solid arrows  $p$  and  $f$  go from  $a$  and  $b$  to  $b$  and  $c$  respectively.)*

com  $f \circ h = f \circ g$ . Então, existe um único  $k$  tal que  $h = p \circ k$  e  $g = q \circ k$ . Como  $p = q$ , segue-se que  $g = h$ , isto é,  $f$  é monomorfismo. □

FATO 12. Se o quadrado dado pelo diagrama

$$\begin{array}{ccc} a & \xrightarrow{f'} & b \\ g' \downarrow & & \downarrow g \\ c & \xrightarrow{f} & d \end{array}$$

é um pullback e  $g$  é monomorfismo, então  $g'$  também é monomorfismo.

**Prova.** Considere os morfismos  $u, v : e \rightarrow a$  tais que  $g' \circ u = g' \circ v$ . Queremos mostrar que  $u = v$ . Faça  $g'' = g' \circ u$  e  $f'' = f' \circ u$  (ver diagrama).

$$\begin{array}{ccccc} & & e & & \\ & & \swarrow & & \searrow \\ & & u, v & & f'' \\ & & \searrow & & \swarrow \\ & & a & \xrightarrow{f'} & b \\ & & \downarrow g' & & \downarrow g \\ & & c & \xrightarrow{f} & d \\ & & \swarrow g'' & & \swarrow f'' \\ & & e & & \end{array}$$

Temos, então, que

$$\begin{aligned} f \circ g'' &= f \circ g' \circ u && \text{pois } g'' = g' \circ u \\ &= g \circ f' \circ u && \text{pois } g \circ f' = f \circ g' \\ &= g \circ f'' && \text{pois } f'' = f' \circ u \end{aligned}$$

Como o quadrado é um pullback,  $u$  é o único morfismo tal que  $g'' = g' \circ u$  e  $f'' = f' \circ u$ . Como  $g' \circ u = g' \circ v$ , temos também que  $g'' = g' \circ v$ . Segue-se, então, que

$$\begin{aligned} g \circ f' \circ v &= f \circ g' \circ v && \text{pois } g \circ f' = f \circ g' \\ &= f \circ g' \circ u && \text{pois } g' \circ u = g' \circ v \\ &= f \circ g'' && \text{pois } g' \circ u = g'' \\ &= g \circ f'' && \text{estabelecido acima} \end{aligned}$$

Como  $g$  é monomorfismo (cancelável à esquerda), temos  $f'' = f' \circ v$ . Assim,  $v$  também fatora ambos  $g''$  e  $f''$ . Como  $u$  é única, temos que  $u = v$ . Logo,  $g'$  é monomorfismo.  $\square$

Voltemos, agora, para o ambiente de uma categoria com morfismos verdade  $\mathcal{E}$ . Os fatos estabelecidos acima implicam o seguinte resultado.

LEMA 13. Em  $\mathcal{E}$ , se o quadrado dado pelo diagrama

$$\begin{array}{ccc} a & \xrightarrow{g} & b \\ g \downarrow & & \downarrow f \\ b & \xrightarrow{f} & c \end{array}$$

é um pullback, então  $a \cong b$ , isto é,  $a$  e  $b$  são objetos isomorfos.

**Prova.** Pelo fato 10,  $g$  é epimorfismo. Pelo fato 11,  $f$  é monomorfismo. Assim, pelo fato 12,  $g$  também é monomorfismo. Como  $\mathcal{E}$  é uma categoria com morfismos verdade, segue-se do corolário 3 que  $g$  é isomorfismo. Logo,  $a \cong b$ .  $\square$

Segue-se, então, uma propriedade que estabelece uma relação entre os valores de verdade verdadeiro e falso.

PROPOSIÇÃO 14. *Em uma categoria com morfismos verdade não degenerada, temos que  $\top \neq \perp$ .*

**Prova.** O morfismo falso é definido pelo pullback

$$\begin{array}{ccc} 0 & \xrightarrow{0_1} & 1 \\ 0_1 \downarrow & & \downarrow \perp \\ 1 & \xrightarrow{\top} & \Omega \end{array}$$

Suponha que  $\top = \perp$ . Pelo lema 13, temos que  $0 \cong 1$ , mas, nesse caso, a categoria é degenerada.  $\square$

Segue-se assim, que se  $\top = \perp$ , então  $\mathcal{E}$  é uma categoria degenerada. (Todos os objetos são isomorfos entre si.)

### 5. Morfismos verdade em CTM's

Considere uma categoria com morfismos verdade  $\mathcal{E}$ . O classificador de subobjetos  $\Omega$  é um  $\mathcal{E}$ -objeto cujos elemento (morfismos do tipo  $1 \rightarrow \Omega$ ) são denominados valores de verdade. Se  $\mathcal{E}$  é não degenerada, temos pelo menos dois elementos em  $\Omega$ , os morfismos  $\top, \perp : 1 \rightarrow \Omega$ . Queremos, agora, definir certos morfismos em  $\mathcal{E}$  que corresponderão às funções de verdade associadas aos conectivos de negação, conjunção, implicação e disjunção das linguagens proposicionais usuais.

Se  $a$  é um  $\mathcal{E}$ -objeto e  $n$  é um inteiro positivo, vamos denotar por  $a^n$  o produto (categorial) de  $n$  cópias de  $a$ , isto é:

$$a^n := \underbrace{a \times a \times \dots \times a}_{n \text{ vezes}}$$

Assim, um *morfismo verdade* em  $\mathcal{E}$  é qualquer morfismo do tipo:

$$\Omega^n \rightarrow \Omega,$$

para algum inteiro positivo  $n$ . Como de costume, o número  $n$  é dito a *aridade* do morfismo verdade. Estamos interessados principalmente nos casos em que  $n$  é igual a 1 ou 2.

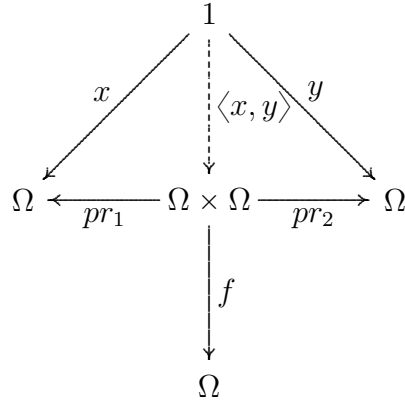
Suponhamos, por exemplo que temos um morfismo binário (de aridade 2)

$$f : \Omega \times \Omega \rightarrow \Omega.$$

Considere, então, dois valores de verdade (elementos de  $\Omega$ ) quaisquer  $x, y : 1 \rightarrow \Omega$ . Tomando o produto  $\langle x, y \rangle$  dado pelo diagrama

$$\begin{array}{ccccc} & & 1 & & \\ & \swarrow & \downarrow & \searrow & \\ & x & \langle x, y \rangle & y & \\ \Omega & \xleftarrow{pr_1} & \Omega \times \Omega & \xrightarrow{pr_2} & \Omega \end{array}$$

e, compondo com o morfismo  $f$  obtemos (ver diagrama)

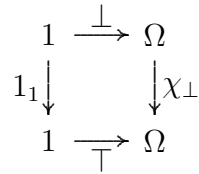


um morfismo  $f \circ \langle x, y \rangle : 1 \rightarrow \Omega$ , que é também um elemento de  $\Omega$  (um valor de verdade).

Assim, dado um morfismo verdade  $f$  qualquer, sempre podemos avaliar o resultado de compor  $f$  com morfismos produto dos valores de verdade  $\top$  e  $\perp$ .

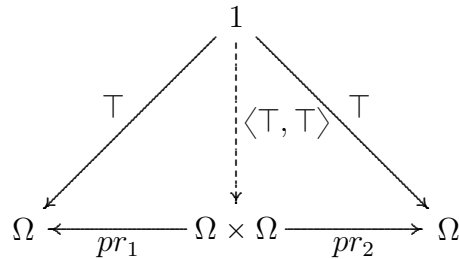
No que se segue, vamos mostrar como construir os morfismos verdade no ambiente das CTM's. O que justifica sua qualificação como categorias com morfismos verdade. Além disso, vamos estudar o comportamento desses morfismos com respeito aos valores de verdade  $\top$  e  $\perp$ . Esse estudo justificará denominar tais morfismos como morfismo negação, conjunção, implicação e disjunção.

**5.1. Morfismo negação.** Como  $1$  é terminal em  $\mathcal{E}$ , os valores de verdade são sempre monomorfismos. Considere, então, o monomorfismo  $\perp : 1 \rightarrow \Omega$  que é o falso. Aplicando o  $\Omega$ -Axioma a  $\perp$ , temos o seguinte pullback:

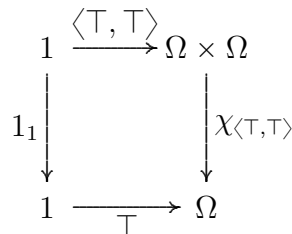


Definimos, então, o morfismo verdade  $f_{\perp} : \Omega \rightarrow \Omega$  como o caráter de  $\perp$ , isto é,  $f_{\perp} := \chi_{\perp}$ . Tal morfismo verdade é dito o *morfismo negação*.

**5.2. Morfismo conjunção.** Considere o produto  $\Omega \times \Omega$  e construa o morfismo produto  $\langle \top, \top \rangle$  dado pelo diagrama:



Novamente, como  $1$  é terminal,  $\langle \top, \top \rangle : 1 \rightarrow \Omega \times \Omega$  é um monomorfismo. Aplicando o  $\Omega$ -Axioma a  $\langle \top, \top \rangle$ , temos o seguinte pullback:





Definimos, então, o morfismo verdade  $f_{\wedge} : \Omega \times \Omega \rightarrow \Omega$  como o caráter de  $\langle \top, \top \rangle$ , isto é,  $f_{\wedge} := \chi_{\langle \top, \top \rangle}$ . Tal morfismo verdade é dito o *morfismo conjunção*.

**5.3. Morfismo implicação.** Considere, agora, o par de morfismos verdade dados por:

$$f_{\wedge}, pr_1 : \Omega \times \Omega \rightarrow \Omega.$$

o morfismo conjunção e a primeira projeção do produto. (Note que as projeções, nesse caso, também são morfismos verdade.)

Tome um equalizador desse par de morfismos.

$$(\leq) \xrightarrow{e} \Omega \times \Omega \xrightarrow{f_{\wedge}, pr_1} \Omega$$

Como todo equalizador é monomorfismo, aplicando o  $\Omega$ -Axioma a  $e : (\leq) \rightarrow \Omega \times \Omega$ , temos o seguinte pullback:

$$\begin{array}{ccc} (\leq) & \xrightarrow{e} & \Omega \times \Omega \\ \downarrow ! & & \downarrow \chi_e \\ 1 & \xrightarrow{\top} & \Omega \end{array}$$

Definimos, então, o morfismo verdade  $f_{\supset} : \Omega \times \Omega \rightarrow \Omega$  como o caráter de  $e$ , um equalizador de  $f_{\wedge}$  e  $pr_1$ , isto é,  $f_{\supset} := \chi_e$ . Tal morfismo verdade é dito o *morfismo implicação*.

**5.4. Morfismo disjunção.** Considere o morfismo verdade  $\top_{\Omega} : \Omega \rightarrow \Omega$  definido como  $\top_{\Omega} := \top \circ |_{\Omega} : \Omega \rightarrow 1 \rightarrow \Omega$ . tome o morfismo produto  $\langle \top_{\Omega}, 1_{\Omega} \rangle : \Omega \rightarrow \Omega \times \Omega$  dado pelo diagrama

$$\begin{array}{ccccc} & & \Omega & & \\ & \swarrow & \vdots & \searrow & \\ & \top_{\Omega} & \langle \top_{\Omega}, 1_{\Omega} \rangle & 1_{\Omega} & \\ & \Omega & \Omega \times \Omega & \Omega & \\ & \xleftarrow{pr_1} & & \xrightarrow{pr_2} & \end{array}$$

Tome, também o produto invertido  $\langle 1_{\Omega}, \top_{\Omega} \rangle : \Omega \rightarrow \Omega \times \Omega$ . Forme, agora, o coproduto

$$[\langle \top_{\Omega}, 1_{\Omega} \rangle, \langle 1_{\Omega}, \top_{\Omega} \rangle] : \Omega + \Omega \rightarrow \Omega \times \Omega$$

dado pelo diagrama:

$$\begin{array}{ccccc} \Omega & \xrightarrow{i_1} & \Omega + \Omega & \xleftarrow{i_2} & \Omega \\ & \searrow & \downarrow f & \swarrow & \\ \langle \top_{\Omega}, 1_{\Omega} \rangle & & \Omega \times \Omega & & \langle 1_{\Omega}, \top_{\Omega} \rangle \end{array}$$

com  $f := [\langle \top_{\Omega}, 1_{\Omega} \rangle, \langle 1_{\Omega}, \top_{\Omega} \rangle]$ .

Aplicando o processo de construção de imagens de morfismos a  $f$ , obtemos um morfismo

$$f(\Omega + \Omega) \xrightarrow{imf} \Omega \times \Omega,$$

(que é um equalizador do pushout de  $f$  por  $f$ ).

Como  $imf$  é monomorfismo, aplicando o  $\Omega$ -Axioma a  $imf : f(\Omega + \Omega) \rightarrow \Omega \times \Omega$ , temos o seguinte pullback:

$$\begin{array}{ccc} f(\Omega + \Omega) & \xrightarrow{imf} & \Omega \times \Omega \\ \downarrow ! & & \downarrow \chi_{imf} \\ 1 & \xrightarrow{\top} & \Omega \end{array}$$

Definimos, então, o morfismo verdade  $f_{\vee} : \Omega \times \Omega \rightarrow \Omega$  como o caráter da imagem do morfismo coproduto  $[\langle \top_{\Omega}, 1_{\Omega} \rangle, \langle 1_{\Omega}, \top_{\Omega} \rangle]$ , isto é,  $f_{\vee} := \chi_f$ . Tal morfismo verdade é dito o *morfismo disjunção*.

**5.5. Resumindo.** Vamos resumir as definições dos morfismos verdade construídos até o momento.

- $f_{\perp} : \Omega \rightarrow \Omega$  é o caráter de  $\perp : 1 \rightarrow \Omega$ ;
- $f_{\wedge} : \Omega \times \Omega \rightarrow \Omega$  é o caráter do morfismo produto  $\langle \top, \top \rangle : 1 \rightarrow \Omega \times \Omega$ ;
- $f_{\supset} : \Omega \times \Omega \rightarrow \Omega$  é o caráter equalizador do par de morfismos  $f_{\wedge}, pr_1 : \Omega \times \Omega \rightarrow \Omega$ ;
- $f_{\vee} : \Omega \times \Omega \rightarrow \Omega$  é o caráter da imagem do morfismo coproduto  $[\langle \top_{\Omega}, 1_{\Omega} \rangle, \langle 1_{\Omega}, \top_{\Omega} \rangle] : \Omega + \Omega \rightarrow \Omega \times \Omega$ .

É importante, nesse ponto, observar que todas estas construções apresentadas podem ser executadas em uma categoria com morfismos verdade. (Tais categorias poderiam ser denominadas *categorias proposicionais*.) Note que, em nenhum passo dessas construções, foi necessário que a categoria tivesse exponenciação. *Não é preciso operar em um ambiente de topos se estamos interessados apenas em linguagens proposicionais.* A única consequência da existência de exponenciação que precisamos é o que denominamos 0-Axioma: *todo morfismo cujo codomínio é um objeto inicial é um isomorfismo*. Segue-se desse axioma que morfismos do tipo  $a \rightarrow 0$  são monomorfismos, e isso foi necessário para definir o valor de verdade *falso*; e em seguida o morfismo negação. Todo o resto é obtido por meio de limites, colimites, imagens e o  $\Omega$ -Axioma, cujas construções estão disponíveis em CTM's.

Nosso próximo passo é estudar o comportamento dos morfismos verdade com respeito aos valores de verdade  $\top$  e  $\perp$ , justificando, assim, a denominação dos mesmos. É esse estudo que faremos agora.

## 6. Morfismos verdade e valores de verdade

Vamos, agora, investigar o comportamento dos morfismos verdade com respeito aos valores de verdade *verdadeiro* e *falso*.

**6.1. Tabela do morfismo negação.** Observe que, da comutatividade do diagrama pullback que define  $f_{\neg}$ ,

$$\begin{array}{ccc} 1 & \xrightarrow{\perp} & \Omega \\ 1_1 \downarrow & & \downarrow f_{\neg} \\ 1 & \xrightarrow{\top} & \Omega \end{array}$$

segue-se que  $f_{\neg} \circ \perp = \top \circ 1_1 = \top$ .

Por outro lado, considere o diagrama:

$$\begin{array}{ccc}
 0 & \xrightarrow{0_1} & 1 \\
 0_1 \downarrow & & \downarrow \top \\
 1 & \xrightarrow{\perp} & \Omega \\
 1_1 \downarrow & & \downarrow f_{\neg} \\
 1 & \xrightarrow{\top} & \Omega
 \end{array}$$

O quadrado superior é o pullback (invertido) que define  $\perp$ . O quadrado inferior é o pullback que define  $f_{\neg}$ . Logo, pelo lema do Pullback, o retângulo externo é um pullback. Assim, pelo  $\Omega$ -Axioma, temos  $f_{\neg} \circ \top = \chi_{0_1} = \perp$ .

Podemos, então, resumir o comportamento (composição) do morfismo negação  $f_{\neg}$  com respeito aos valores de verdade  $\top, \perp : 1 \rightarrow \Omega$  na seguinte tabela:

$x$	$f_{\neg} \circ x$
$\top$	$\perp$
$\perp$	$\top$

que corresponde à tabela usual da negação da lógica proposicional clássica.

**6.2. Tabela do morfismo conjunção.** Novamente, pela comutatividade do diagrama

$$\begin{array}{ccc}
 1 & \xrightarrow{\langle \top, \top \rangle} & \Omega \times \Omega \\
 1_1 \downarrow & & \downarrow f_{\wedge} \\
 1 & \xrightarrow{\top} & \Omega
 \end{array}$$

segue-se que  $f_{\wedge} \circ \langle \top, \top \rangle = \top \circ 1_1 = \top$ .

Considere, agora, o seguinte diagrama:

$$\begin{array}{ccc}
 0 & \xrightarrow{0_1} & 1 \\
 0_1 \downarrow & & \downarrow x \\
 1 & \xrightarrow{\langle \top, \top \rangle} & \Omega \times \Omega \\
 1_1 \downarrow & & \downarrow f_{\wedge} \\
 1 & \xrightarrow{\top} & \Omega
 \end{array}$$

em que  $x$  é  $\langle \top, \perp \rangle$ ,  $\langle \perp, \top \rangle$  ou  $\langle \perp, \perp \rangle$ .

O quadrado inferior do diagrama acima é o pullback que define  $f_{\wedge}$ . Vamos mostrar que o quadrado superior é um pullback para qualquer das escolhas de  $x$ . Seguirá daí, pelo Lema do Pullback, que o retângulo externo é um pullback. Então, pelo  $\Omega$ -Axioma, teremos que  $f_{\wedge} \circ x = \chi_{0_1} = \perp$ .

Considere, então, o diagrama superior

$$\begin{array}{ccc}
 0 & \xrightarrow{0_1} & 1 \\
 0_1 \downarrow & & \downarrow x \\
 1 & \xrightarrow{\langle \top, \top \rangle} & \Omega \times \Omega
 \end{array}$$

Como 0 é inicial, o quadrado é comutativo. Seja, então,  $g : a \rightarrow 1$  tal que  $\langle \top, \top \rangle \circ g = x \circ g$ .

$$\begin{array}{ccc}
 a & & \\
 \swarrow g & & \searrow g \\
 & \begin{array}{ccc}
 0 & \xrightarrow{0_1} & 1 \\
 0_1 \downarrow & & \downarrow x \\
 1 & \xrightarrow{\langle \top, \top \rangle} & \Omega \times \Omega
 \end{array} & & 
 \end{array}$$

Temos três casos:

- (1) Se  $x = \langle \top, \perp \rangle$ , então  $\langle \top, \top \rangle \circ g = \langle \top, \perp \rangle \circ g$  e temos  $\langle \top \circ g, \top \circ g \rangle = \langle \top \circ g, \perp \circ g \rangle$ , obtendo  $\top \circ g = \perp \circ g$ .
- (2) Se  $x = \langle \perp, \top \rangle$ , então um raciocínio análogo resulta que  $\top \circ g = \perp \circ g$ .
- (3) Se  $x = \langle \perp, \perp \rangle$ , então também segue que  $\top \circ g = \perp \circ g$ .

Portanto, em qualquer caso temos que  $\top \circ g = \perp \circ g$ . Tome, agora, o pullback que define o valor de verdade  $\perp$ .

$$\begin{array}{ccc}
 a & & \\
 \swarrow h & & \searrow g \\
 & \begin{array}{ccc}
 0 & \xrightarrow{0_1} & 1 \\
 0_1 \downarrow & & \downarrow \perp \\
 1 & \xrightarrow{\top} & \Omega
 \end{array} & & 
 \end{array}$$

Como o quadrado é um pullback e  $\top \circ g = \perp \circ g$ , segue-se que existe um único  $h : a \rightarrow 0$ . (A comutatividade dos triângulos é óbvia, pois 1 é terminal em  $\mathcal{E}$ .) Pelo 0-Axioma, tem-se que  $h$  é iso e, assim,  $a$  é inicial em  $\mathcal{E}$ . Transortando  $h$  para nosso diagrama inicial temos:

$$\begin{array}{ccc}
 a & & \\
 \swarrow h & & \searrow g \\
 & \begin{array}{ccc}
 0 & \xrightarrow{0_1} & 1 \\
 0_1 \downarrow & & \downarrow x \\
 1 & \xrightarrow{\langle \top, \top \rangle} & \Omega \times \Omega
 \end{array} & & 
 \end{array}$$

Como 1 é terminal,  $0_1 \circ h = g$  e, assim, o quadrado é um pullback. Segue-se, então, que

$$f_{\wedge} \circ \langle \top, \perp \rangle = f_{\wedge} \circ \langle \perp, \top \rangle = f_{\wedge} \circ \langle \perp, \perp \rangle = \perp.$$

Podemos, então, resumir o comportamento (composição) do morfismo conjunção  $f_{\wedge}$  com respeito aos valores de verdade  $\top, \perp : 1 \rightarrow \Omega$  na seguinte tabela:

$x$	$y$	$f_{\wedge} \circ \langle x, y \rangle$
$\top$	$\top$	$\top$
$\top$	$\perp$	$\perp$
$\perp$	$\top$	$\perp$
$\perp$	$\perp$	$\perp$

que corresponde à tabela usual da conjunção da lógica proposicional clássica.

**6.3. Tabela do morfismo implicação.** Considere o seguinte diagrama:

$$\begin{array}{ccc}
 & & 1 \\
 & \swarrow h & \downarrow x \\
 (\leq) & \xrightarrow{e} & \Omega \times \Omega \xrightarrow{f_{\wedge}, pr_1} \Omega \\
 \downarrow ! & & \downarrow f_{\supset} \\
 1 & \xrightarrow{\top} & \Omega
 \end{array}$$

em que  $x$  é  $\langle \top, \top \rangle$ ,  $\langle \perp, \top \rangle$  ou  $\langle \perp, \perp \rangle$ . Nesse caso, para qualquer escolha de  $x$ , usando a tabela para o morfismo conjunção e as propriedades da projeção  $pr_1$ , temos que:

$$f_{\wedge} \circ x = pr_1 \circ x.$$

Como  $e$  é um equalizador de  $f_{\wedge}$  e  $pr_1$ , existe um único morfismo  $h : 1 \rightarrow (\leq)$  tal que  $e \circ h = x$ . Mas, então, como o quadrado é o pullback que define  $f_{\supset}$ , temos, por comutatividade, que  $\top \circ ! = f_{\supset} \circ e$ . Assim,

$$\begin{aligned}
 f_{\supset} \circ x &= f_{\supset} \circ e \circ h \\
 &= \top \circ ! \circ h \\
 &= \top \circ 1_1 \\
 &= \top
 \end{aligned}$$

Segue-se, então, que:

$$f_{\supset} \circ \langle \top, \top \rangle = f_{\supset} \circ \langle \perp, \top \rangle = f_{\supset} \circ \langle \perp, \perp \rangle = \top.$$

Resta calcular o valor de verdade de  $f_{\supset}$  aplicado ao morfismo produto  $\langle \top, \perp \rangle$ . Para isso, considere o diagrama:

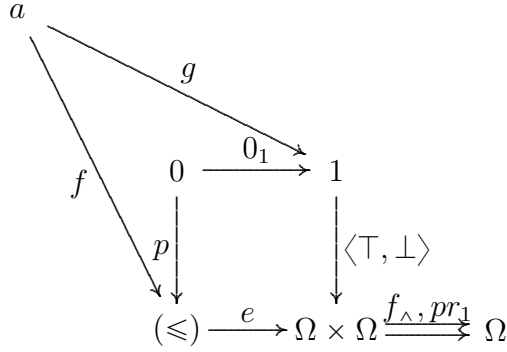
$$\begin{array}{ccc}
 0 & \xrightarrow{0_1} & 1 \\
 p \downarrow & & \downarrow \langle \top, \perp \rangle \\
 (\leq) & \xrightarrow{e} & \Omega \times \Omega \xrightarrow{f_{\wedge}, pr_1} \Omega \\
 \downarrow ! & & \downarrow f_{\supset} \\
 1 & \xrightarrow{\top} & \Omega
 \end{array}$$

O quadrado inferior é o pullback que define  $f_{\supset}$ . Mostremos que o quadrado superior é um pullback. Seguirá daí que o retângulo externo é um pullback e, pelo  $\Omega$ -Axioma, teremos que

$$f_{\supset} \circ \langle \top, \perp \rangle = \chi_{o_1} = \perp,$$

obtendo o que desejamos.

Examinemos, então, o quadrado superior:



Como  $0$  é inicial, o quadrado é comutativo. Considere os morfismos  $f : a \rightarrow (\leq)$  e  $g : a \rightarrow 1$ . Suponha que  $e \circ f = \langle \top, \perp \rangle \circ g$ . Como  $e$  equaliza  $f_{\wedge}$  e  $pr_1$ , temos que  $f_{\wedge} \circ e = pr_1 \circ e$ . Compondo com  $f$  temos:

$$f_{\wedge} \circ e \circ f = pr_1 \circ e \circ f.$$

Mas, por um lado,

$$f_{\wedge} \circ e \circ f = f_{\wedge} \circ \langle \top, \perp \rangle \circ g = \perp \circ g.$$

(A última igualdade segue da aplicação da tabela para o morfismo conjunção.)

Por outro lado,

$$pr_1 \circ e \circ f = pr_1 \circ \langle \top, \perp \rangle \circ g = \top \circ g$$

(Aplicando a propriedade da projeção.)

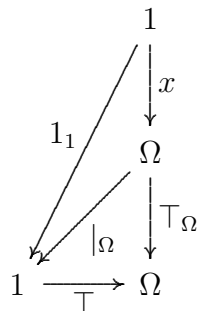
Portanto, temos que  $\top \circ g = \perp \circ g$ . Agora, por um argumento análogo ao caso do morfismo conjunção, temos que o quadrado acima é um pullback. Logo, segue-se que  $f_{\supset} \circ \langle \top, \perp \rangle = \perp$ , como queríamos.

Podemos, então, resumir o comportamento (composição) do morfismo implicação  $f_{\supset}$  com respeito aos valores de verdade  $\top, \perp : 1 \rightarrow \Omega$  na seguinte tabela:

$x$	$y$	$f_{\supset} \circ \langle x, y \rangle$
$\top$	$\top$	$\top$
$\top$	$\perp$	$\perp$
$\perp$	$\top$	$\top$
$\perp$	$\perp$	$\top$

que corresponde à tabela usual da implicação material da lógica proposicional clássica.

**6.4. Tabela do morfismo disjunção.** Observemos, primeiramente que, do diagrama



temos:  $\top_\Omega \circ x = \top \circ |_\Omega \circ x = \top \circ 1_1 = \top$ . (Assim, a composição de  $\top_\Omega$  com qualquer valor de verdade resulta em  $\top$ .)

Segue-se, então, que  $\top_\Omega \circ \top = \top_\Omega \circ \perp = \top$ .

O processo de construção de  $\text{im}[\langle \top_\Omega, 1_\Omega \rangle, \langle 1_\Omega, \top_\Omega \rangle]$  provém do pushout dado por:

$$\begin{array}{ccc} \Omega + \Omega & \xrightarrow{f} & \Omega \times \Omega \\ f \downarrow & & \downarrow q \\ \Omega \times \Omega & \xrightarrow{p} & a \end{array}$$

em que  $f := [\langle \top_\Omega, 1_\Omega \rangle, \langle 1_\Omega, \top_\Omega \rangle]$ .

Então, temos que:

$$p \circ [\langle \top_\Omega, 1_\Omega \rangle, \langle 1_\Omega, \top_\Omega \rangle] = q \circ [\langle \top_\Omega, 1_\Omega \rangle, \langle 1_\Omega, \top_\Omega \rangle],$$

Isto é,

$$[p \circ \langle \top_\Omega, 1_\Omega \rangle, p \circ \langle 1_\Omega, \top_\Omega \rangle] = [q \circ \langle \top_\Omega, 1_\Omega \rangle, q \circ \langle 1_\Omega, \top_\Omega \rangle].$$

Temos, então, duas equações:

$$\begin{aligned} p \circ \langle \top_\Omega, 1_\Omega \rangle &= q \circ \langle \top_\Omega, 1_\Omega \rangle & (I) \\ p \circ \langle 1_\Omega, \top_\Omega \rangle &= q \circ \langle 1_\Omega, \top_\Omega \rangle & (II) \end{aligned}$$

Compondo a equação (I) com  $\top$ , e usando a propriedade do início da seção, temos:

$$\begin{aligned} p \circ \langle \top_\Omega, 1_\Omega \rangle \circ \top &= q \circ \langle \top_\Omega, 1_\Omega \rangle \circ \top \\ p \circ \langle \top_\Omega \circ \top, 1_\Omega \circ \top \rangle &= q \circ \langle \top_\Omega \circ \top, 1_\Omega \circ \top \rangle \\ p \circ \langle \top, \top \rangle &= q \circ \langle \top, \top \rangle \end{aligned}$$

Compondo a equação (I) com  $\perp$  temos:

$$\begin{aligned} p \circ \langle \top_\Omega, 1_\Omega \rangle \circ \perp &= q \circ \langle \top_\Omega, 1_\Omega \rangle \circ \perp \\ p \circ \langle \top_\Omega \circ \perp, 1_\Omega \circ \perp \rangle &= q \circ \langle \top_\Omega \circ \perp, 1_\Omega \circ \perp \rangle \\ p \circ \langle \top, \perp \rangle &= q \circ \langle \top, \perp \rangle \end{aligned}$$

Compondo a equação (II) com  $\perp$  temos:

$$\begin{aligned} p \circ \langle 1_\Omega, \top_\Omega \rangle \circ \perp &= q \circ \langle 1_\Omega, \top_\Omega \rangle \circ \perp \\ p \circ \langle 1_\Omega \circ \perp, \top_\Omega \circ \perp \rangle &= q \circ \langle 1_\Omega \circ \perp, \top_\Omega \circ \perp \rangle \\ p \circ \langle \perp, \top \rangle &= q \circ \langle \perp, \top \rangle \end{aligned}$$

Assim, se  $x : 1 \rightarrow \Omega \times \Omega$  for  $\langle \top, \top \rangle$ ,  $\langle \top, \perp \rangle$  ou  $\langle \perp, \top \rangle$ , segue-se que  $p \circ x = q \circ x$ . Considere, então, o seguinte diagrama em que  $f := [\langle \top_\Omega, 1_\Omega \rangle, \langle 1_\Omega, \top_\Omega \rangle]$  e  $x$  é como acima:

$$\begin{array}{ccccc} & & 1 & & \\ & & \downarrow x & & \\ & \swarrow h & & & \\ f(\Omega + \Omega) & \xrightarrow{\text{im}f} & \Omega \times \Omega & \xrightarrow{p, q} & a \\ \downarrow ! & & \downarrow f_\vee & & \\ 1 & \xrightarrow{\top} & \Omega & & \end{array}$$

Como  $imf$  é equalizador de  $p$  e  $q$  e, além disso,  $p \circ x = q \circ x$ , segue-se que existe um único  $h : 1 \rightarrow f(\Omega + \Omega)$  tal que  $imf \circ h = x$ . Como o quadrado comuta, temos

$$f_v \circ imf = \top \circ !$$

Compondo com  $h$ , temos:

$$f_v \circ \underbrace{imf \circ h}_x = \top \circ \underbrace{! \circ h}_{1_1}.$$

Portanto,  $f_v \circ x = \top$ .

Assim, segue-se que  $f_v \circ \langle \top, \top \rangle = f_v \circ \langle \perp, \top \rangle = f_v \circ \langle \perp, \perp \rangle = \top$ .

Até o momento, o comportamento do morfismo  $f_v$  com respeito aos valores de verdade podem ser resumidos de acordo com a seguinte tabela:

$x$	$y$	$f_v \circ \langle x, y \rangle$
$\top$	$\top$	$\top$
$\top$	$\perp$	$\top$
$\perp$	$\top$	$\top$
$\perp$	$\perp$	?

Resta mostrar que  $f_v \circ \langle \perp, \perp \rangle = \perp$ . Mas, nesse ponto, parece que temos um problema. Seguindo os métodos de prova dos casos anteriores, parece que a estratégia a ser adotada é a seguinte:

Consideremos o diagrama

$$\begin{array}{ccc}
 0 & \xrightarrow{0_1} & 1 \\
 \downarrow !r & & \downarrow \langle \perp, \perp \rangle \\
 f(\Omega + \Omega) & \xrightarrow{imf} & \Omega \times \Omega \\
 \downarrow ! & & \downarrow f_v \\
 1 & \xrightarrow{\top} & \Omega
 \end{array}$$

Como o quadrado inferior é o pullback que define  $f_v$ , se mostrarmos que o quadrado superior é um pullback, então, pelo Lema do Pullback, o retângulo exterior é um pullback e, pelo  $\Omega$ -Axioma, temos que  $f_v \circ \langle \perp, \perp \rangle = \chi_{0_1} = \perp$ , estabelecendo o caso que falta.

Assim, nossa única tarefa é mostrar que o quadrado

$$\begin{array}{ccc}
 0 & \xrightarrow{0_1} & 1 \\
 \downarrow r & & \downarrow \langle \perp, \perp \rangle \\
 f(\Omega + \Omega) & \xrightarrow{imf} & \Omega \times \Omega
 \end{array}$$

é um pullback.

Ocorre que, aparentemente, a prova desse fato necessita a introdução de hipóteses adicionais na nossa CTM  $\mathcal{E}$ . Essas hipóteses serão apresentadas na seção a seguir, definindo um tipo de CTM que denominaremos *prototopos*. O caso que resta para ser demonstrado será chamado de *caso anômalo*.



## 7. Prototopos

A fim de lidar com o caso anômalo, vamos definir um tipo de categoria com morfismos verdade com algumas propriedades adicionais.

Dizemos que uma categoria  $\mathcal{E}$  é um *prototopos* se e somente se  $\mathcal{E}$  satisfaz os seguintes axiomas:

(PT1)  $\mathcal{E}$  é uma categoria com morfismos verdade. Isto é,  $\mathcal{E}$  é finitamente bicompleta, possui classificador de subobjetos e satisfaz o 0-Axioma.

(PT2) Em  $\mathcal{E}$ , pullbacks preservam epimorfismos. Isto é, se o quadrado

$$\begin{array}{ccc} a & \xrightarrow{f'} & b \\ g' \downarrow & & \downarrow g \\ c & \xrightarrow{f} & d \end{array}$$

é um pullback e  $f$  é um epimorfismo, então  $f'$  também é epimorfismo.

(PT3) Em  $\mathcal{E}$ , coprodutos preservam pullbacks. Isto é, se os diagramas abaixo

$$\begin{array}{ccc} a & \xrightarrow{f} & d \\ g \downarrow & & \downarrow k \\ b & \xrightarrow{h} & e \end{array} \quad \begin{array}{ccc} a' & \xrightarrow{f'} & d \\ g' \downarrow & & \downarrow k \\ b' & \xrightarrow{h'} & e \end{array}$$

são pullbacks, então o quadrado

$$\begin{array}{ccc} a + a' & \xrightarrow{[f, f']} & d \\ g + g' \downarrow & & \downarrow k \\ b + b' & \xrightarrow{[h, h']} & e \end{array}$$

é um pullback também.

É preciso mencionar que (PT2) e (PT3) são propriedades que valem em topos e são consequências do Teorema Fundamental da Teoria de Topos: *Se  $\mathcal{E}$  é um topos e  $a$  é um  $\mathcal{E}$ -objeto, então a categoria slice  $\mathcal{E}/a$  é também um topos.*

Dessas propriedades, agora no ambiente de prototopos, segue um resultado que servirá de base para a demonstração do caso anômalo.

LEMA 15. *Em qualquer prototopos  $\mathcal{E}$ , se o quadrado*

$$\begin{array}{ccc} a & \xrightarrow{f} & b \\ u \downarrow & & \downarrow v \\ c & \xrightarrow{g} & d \end{array}$$

*é um pullback, então existe um morfismo*

$$h : f(a) \rightarrow g(c)$$

que torna o quadrado à direita do diagrama

$$\begin{array}{ccccc}
 a & \xrightarrow{f^*} & f(a) & \xrightarrow{\text{im}f} & b \\
 \downarrow u & & \downarrow h & & \downarrow v \\
 c & \xrightarrow{g^*} & g(c) & \xrightarrow{\text{img}} & d
 \end{array}$$

um pullback<sup>7</sup>.

**Prova.** Considere o diagrama

$$g(c) \xrightarrow{\text{img}} d \xleftarrow{v} b.$$

e construa um pullback do mesmo.

$$\begin{array}{ccc}
 e & \xrightarrow{i} & b \\
 \downarrow h' & & \downarrow v \\
 g(c) & \xrightarrow{\text{img}} & d
 \end{array} \quad (A)$$

Considere, agora, o diagrama

$$\begin{array}{ccccc}
 a & & & & \\
 \swarrow f' & & & & \\
 & e & \xrightarrow{i} & b & \\
 \downarrow g^* \circ u & \downarrow h' & & \downarrow v & \\
 & g(c) & \xrightarrow{\text{img}} & d & 
 \end{array}$$

Como  $\text{img} \circ g^* = g$  (fatoração epi-mono de  $g$ ) e  $g \circ u = f \circ v$  (hipótese), segue-se que  $\text{img} \circ g^* \circ u = f \circ v$  (ou seja, o diagrama externo comuta). Como (A) é um pullback, então existe um único morfismo  $f' : a \rightarrow e$  tal que  $i \circ f' = f$ . Temos, então, o seguinte diagrama comutativo:

$$\begin{array}{ccccc}
 a & \xrightarrow{f'} & e & \xrightarrow{i} & b \\
 \downarrow u & & \downarrow h' & & \downarrow v \\
 c & \xrightarrow{g^*} & g(c) & \xrightarrow{\text{img}} & d
 \end{array}$$

Como, por hipótese, o retângulo exterior é um pullback (pois,  $i \circ f' = f$ ) e o diagrama da direita também é um pullback (diagrama (A)), segue-se, pelo Lema do Pullback, que o quadrado da esquerda é um pullback.

Como  $g^*$  é um epimorfismo, segue-se de (PT2) que  $f'$  é epimorfismo. Assim,

$$f = i \circ f'$$

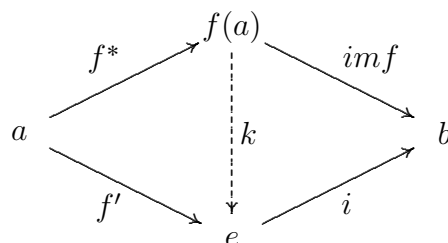
<sup>7</sup> Lembremos que  $\text{im}f \circ f^*$  e  $\text{img} \circ g^*$  são decomposições epi-mono de  $f$  e  $g$ , respectivamente; cuja existência está garantida pelo corolário 8.

é uma fatoração epi-mono de  $f$ . ( $i$  é monomorfismo porque  $img$  é mono e, em qualquer categoria, pullbacks preservam monomorfismos.)

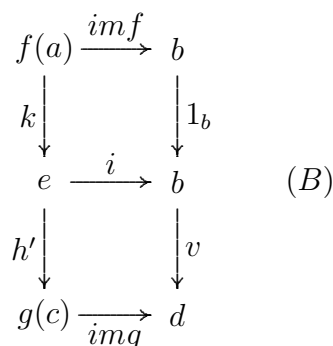
Assim, pelo corolário 8, existe um único isomorfismo

$$k : f(a) \cong e$$

tal que o diagrama a seguir é comutativo:

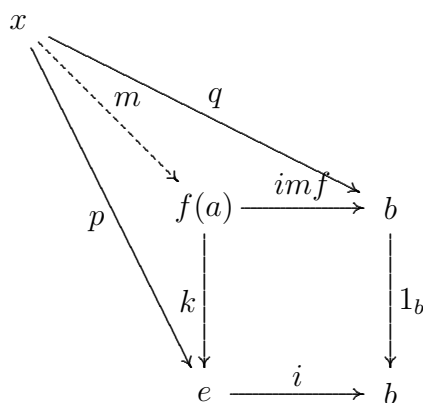


Definindo  $h := h' \circ k$ , temos o seguinte diagrama comutativo:



O quadrado inferior do diagrama acima é um pullback (diagrama (A)). Mostremos que o quadrado superior também é um pullback.

Considere, então, o quadrado



Já sabemos que  $i \circ k = img = 1_b \circ img$ .

Sejam os morfismos  $p : x \rightarrow e$  e  $q : x \rightarrow b$  tais que  $i \circ p = 1_b \circ q = q$ .

Seja  $m := k^{-1} \circ p$  (lembre que  $k$  é iso).

Então, é claro que

$$k \circ m = k \circ k^{-1} \circ p = 1_e \circ p = p.$$

Por outro lado,

$$\begin{aligned} img \circ m &= img \circ k^{-1} \circ p \\ &= i \circ k \circ k^{-1} \circ p \\ &= i \circ 1_e \circ p \\ &= i \circ p \\ &= q \end{aligned}$$

Assim, temos que  $m$  fatora  $p$  e  $q$ . Para ver que  $m$  é único, suponha que exista  $m' : x \rightarrow f(a)$  tal que  $k \circ m' = p$ . Então,  $k \circ m' = p = k \circ m$ . Como  $k$  é mono (pois é iso), segue-se que  $m' = m$ .

Como os dois quadrados de (B) são pullbacks, pelo Lema do Pullback, o retângulo é também um pullback e o lema está demonstrado.  $\square$

### 8. O caso anômalo

PROPOSIÇÃO 16. Em um prototopos  $\mathcal{E}$ , o quadrado dado pelo diagrama

$$\begin{array}{ccc} 0 & \xrightarrow{0_1} & 1 \\ r \downarrow & & \downarrow \langle \perp, \perp \rangle \\ f(\Omega + \Omega) & \xrightarrow{im f} & \Omega \times \Omega \end{array}$$

é um pullback.

**Prova.** Lembrando que  $f$  é dado pelo coproduto

$$f := [\langle \top_\Omega, 1_\Omega \rangle, \langle 1_\Omega, \top_\Omega \rangle] : \Omega + \Omega \rightarrow \Omega \times \Omega$$

e  $im f$  é o equalizador do pushout de  $f$  por  $f$

$$\begin{array}{ccc} \Omega + \Omega & \xrightarrow{f} & \Omega \times \Omega \\ f \downarrow & & \downarrow p \\ \Omega \times \Omega & \xrightarrow{q} & a \end{array}$$

que é dado por

$$f(\Omega + \Omega) \xrightarrow{im f} \Omega \times \Omega \xrightarrow{p, q} a.$$

Considere, então, o seguinte diagrama

$$\begin{array}{ccccc} 0 & \xrightarrow{0_1} & 1 & \xleftarrow{0_1} & 0 \\ \downarrow ! & & \downarrow \langle \perp, \perp \rangle & & \downarrow ! \\ \Omega & \xrightarrow{\langle \top_\Omega, 1_\Omega \rangle} & \Omega \times \Omega & \xleftarrow{\langle 1_\Omega, \top_\Omega \rangle} & \Omega \end{array} \quad (D)$$

Vamos mostrar que ambos os quadrados são pullbacks. O argumento é análogo para ambos os casos. Então, consideremos o diagrama da esquerda.

$$\begin{array}{ccc}
 x & & \\
 \swarrow u & & \searrow \\
 & 0 & \xrightarrow{0_1} & 1 \\
 \downarrow v & \downarrow ! & & \downarrow \langle \perp, \perp \rangle \\
 & \Omega & \xrightarrow{\langle \top_\Omega, 1_\Omega \rangle} & \Omega \times \Omega
 \end{array} \quad (A)$$

Considere os morfismos  $v : x \rightarrow \Omega$  e  $u : x \rightarrow 1$  tais que

$$\langle \top_\Omega, 1_\Omega \rangle \circ v = \langle \perp, \perp \rangle \circ u. \quad (I)$$

Tome o pullback que define  $\perp$ .

$$\begin{array}{ccccc}
 x & & & & \\
 \downarrow v & \swarrow h & \searrow u & & \\
 & 0 & \xrightarrow{0_1} & 1 & \\
 & \downarrow 0_1 & & \downarrow \perp & \\
 \Omega & \xrightarrow{|\Omega} & 1 & \xrightarrow{\top} & \Omega
 \end{array} \quad (B)$$

De (I) segue-se que:

$$\langle \top_\Omega \circ v, 1_\Omega \circ v \rangle = \langle \perp \circ u, \perp \circ u \rangle,$$

e isso nos fornece duas equações:

$$\top_\Omega \circ v = \perp \circ u \quad (IIa)$$

$$v = \perp \circ u \quad (IIb)$$

Pelo corolário 3, temos que  $\perp \circ u = \top_\Omega \circ v = \top \circ |\Omega \circ v$ . Assim, como o quadrado de (B) é um pullback, existe um único  $h : x \rightarrow 0$  tal que  $0_1 \circ h = |\Omega \circ v$  e  $0_1 \circ h = u$ .

Pelo 0-Axioma,  $x$  é inicial em  $\mathcal{E}$  ( $h$  é iso e único).

Transportando  $h$  para o diagrama (A) e, como  $x$  é inicial,  $0_1 \circ h = u$  (já sabíamos) e  $! \circ h = v$ . Logo, o quadrado de (A) é um pullback.

Com o mesmo argumento, temos que o quadrado à direita do diagrama (D) é um pullback também.

Usando, agora, o axioma (PT3) dos prototopos, segue-se que o seguinte diagrama é um pullback

$$\begin{array}{ccc}
 0 + 0 & \xrightarrow{g} & 1 \\
 \downarrow ! & & \downarrow \langle \perp, \perp \rangle \\
 \Omega + \Omega & \xrightarrow{f} & \Omega \times \Omega
 \end{array}$$

em que  $g = [0, 0]$  e  $f = [\langle \top_\Omega, 1_\Omega \rangle, \langle 1_\Omega, \top_\Omega \rangle]$ .

Aplicando o lema 15, segue-se que existe  $r : g(0 + 0) \rightarrow f(\Omega + \Omega)$  que torna o quadrado da direita do diagrama

$$\begin{array}{ccccc}
 0 + 0 & \xrightarrow{g^*} & g(0 + 0) & \xrightarrow{img} & 1 \\
 \downarrow ! & & \downarrow r & & \downarrow \langle \perp, \perp \rangle \\
 \Omega + \Omega & \xrightarrow{f^*} & f(\Omega + \Omega) & \xrightarrow{imf} & \Omega \times \Omega
 \end{array}$$

um pullback.

Mas,  $0 + 0 \cong 0$  e  $g(0 + 0) \cong g(o) \cong 0$ . Assim, temos que o quadrado

$$\begin{array}{ccc}
 0 & \xrightarrow{0_1} & 1 \\
 \downarrow r & & \downarrow \langle \perp, \perp \rangle \\
 f(\Omega + \Omega) & \xrightarrow{imf} & \Omega \times \Omega
 \end{array}$$

é um pullback. Como queríamos demonstrar. □

O raciocínio acima resolve o caso anômalo estabelecendo que  $f_{\vee} \circ \langle \perp, \perp \rangle = \perp$ . Assim, em prototopos, o morfismo disjunção comporta-se, com respeito aos valores de verdade  $\top$  e  $\perp$ , de acordo com a tabela

$x$	$y$	$f_{\vee} \circ \langle x, y \rangle$
$\top$	$\top$	$\top$
$\top$	$\perp$	$\top$
$\perp$	$\top$	$\top$
$\perp$	$\perp$	$\perp$

que corresponde à tabela usual da disjunção na lógica proposicional clássica.

Segue-se, então, uma conclusão geral: Em categorias com morfismos verdade, podem ser definidos todos os morfismos que correspondem às funções de verdade da negação, conjunção, implicação material e disjunção das linguagens proposicionais, fornecendo assim uma semântica categorial para as mesmas. Além disso, em CTM's, os morfismos verdade comportam-se com relação aos valores de verdade exatamente como se comportam as funções clássicas, com uma única exceção correspondente ao que denominamos de caso anômalo. Para este, introduzimos a noção de prototopos que possui a propriedade desejada.

Resta investigar qual é a lógica subjacente à semântica de prototopos. É um resultado conhecido que no caso de topos, a lógica subjacente é o cálculo proposicional intuicionista<sup>8</sup>. Como todo topos é também um prototopos, segue-se que na semântica de prototopos o princípio do terceiro excluído também não vale. Mas, ainda não está totalmente claro que a lógica proposicional intuicionista caracteriza essa semântica. Isso será investigado em trabalhos futuros.

Para completar o presente trabalho vamos mostrar que em CTM's vale um resultado análogo ao teorema fundamental da teoria de topos.

<sup>8</sup> Ver Boileau & Joyal [3].

### 9. O teorema fundamental para CTM

Considere uma categoria  $\mathcal{C}$  qualquer e  $X$  um  $\mathcal{C}$ -objeto. Definimos a categoria slice de  $\mathcal{C}$  por  $X$ , denotada por  $\mathcal{C}/X$ , de forma que:

- $\mathcal{C}/X$ -objetos: são os  $\mathcal{C}$ -morfismos da forma  $A \xrightarrow{a} X$ ;
- $\mathcal{C}/X$ -morfismos: dados os  $\mathcal{C}/X$ -objetos  $A \xrightarrow{a} X$  e  $B \xrightarrow{b} X$ , um  $\mathcal{C}/X$ -morfismo  $f : a \rightarrow b$  é simplesmente um  $\mathcal{C}$ -morfismo  $f : A \rightarrow B$  tal que o seguinte diagrama comuta

$$\begin{array}{ccc} A & \xrightarrow{f} & B \\ & \searrow a & \swarrow b \\ & & X \end{array}$$

isto é,  $b \circ f = a$ .

Composição de morfismos em  $\mathcal{C}/X$  é a composição usual de morfismos em  $\mathcal{C}$  e, também, identidades em  $\mathcal{C}/X$  são identidades em  $\mathcal{C}$ .

Vamos demonstrar, para as CTM's, um teorema análogo ao teorema fundamental da teoria de topós<sup>9</sup>.

**THEOREM 17.** *Se  $\mathcal{E}$  é uma categoria com morfismos verdade e  $X$  é um  $\mathcal{E}$ -objeto, então a categoria slice  $\mathcal{E}/X$  é também uma categoria com morfismos verdade.*

**Prova.** É preciso fazer várias verificações. Lembremos primeiramente que uma condição necessária e suficiente para uma categoria seja finitamente completa é que ela possua objetos terminais, produtos de pares de objetos e equalizadores de pares de morfismos. O mesmo vale para o caso dual, da cocompletude finita.

(I) *O morfismo identidade  $1_X : X \rightarrow X$  é terminal em  $\mathcal{E}/X$ .* De fato, dado um  $\mathcal{E}/X$ -objeto  $A \xrightarrow{a} X$ , esse morfismo é o único  $\mathcal{E}/X$ -morfismo que faz o diagrama abaixo comutar.

$$\begin{array}{ccc} A & \xrightarrow{a} & X \\ & \searrow a & \swarrow 1_X \\ & & X \end{array}$$

Assim,  $a$  é o único morfismo de  $a$  em  $1_X$ , isto é,  $1_X$  é terminal em  $\mathcal{E}/X$ .

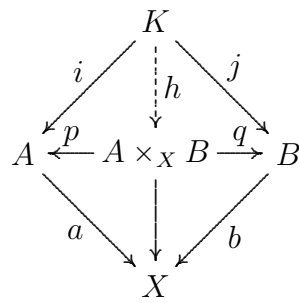
(II) Considere dois  $\mathcal{E}/X$ -objetos  $A \xrightarrow{a} X$  e  $B \xrightarrow{b} X$ . Como  $\mathcal{E}$  é finitamente completa, construa o pullback

$$\begin{array}{ccc} A \times_X B & \xrightarrow{q} & B \\ \downarrow p & & \downarrow b \\ A & \xrightarrow{a} & X \end{array}$$

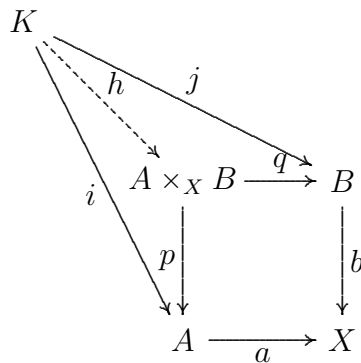
Afirmção: *O  $\mathcal{E}$ -objeto  $a \circ p : A \times_X B \rightarrow X$  junto com os  $\mathcal{E}$ -morfismos  $p : A \times_X B \rightarrow A$  e  $q : A \times_X B \rightarrow B$  formam um produto de  $a$  e  $b$  em  $\mathcal{E}/X$ .*

<sup>9</sup> Ver Freyd [5] e Kock & Wraith [8].

De fato, para ver que temos um produto, considere o diagrama

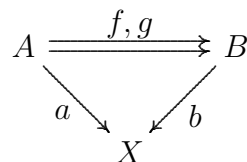


e suponha que  $a \circ i = b \circ j$  (condição para  $i$  e  $j$  serem morfismos em  $\mathcal{E}/X$ ). Transportando para o diagrama do pullback em  $\mathcal{E}$  (ver diagrama)



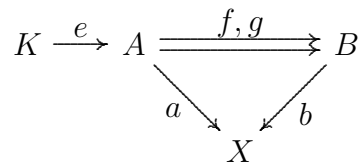
temos que existe um único morfismo  $h : K \rightarrow A \times_X B$  tal que  $p \circ h = i$  e  $q \circ h = j$ . E essa é a condição necessária para a existência de produto em  $\mathcal{E}/X$ . Assim, *um produto em  $\mathcal{E}/X$  é apenas um pullback em  $\mathcal{E}$ .*

(III) Considere dois  $\mathcal{E}/X$ -objetos  $A \xrightarrow{a} X$  e  $B \xrightarrow{b} X$  e dois  $\mathcal{E}/X$ -morfismos paralelos  $f, g : A \rightarrow B$  de forma que o seguinte diagrama comuta:



isto é,  $b \circ f = a = b \circ g$ .

Como  $\mathcal{E}$  é finitamente completa, tome um equalizador de  $f$  e  $g$  em  $\mathcal{E}$ .



Afirmção: *O  $\mathcal{E}/X$ -morfismo  $e : K \rightarrow A$  cujo domínio é o  $\mathcal{E}/X$ -objeto  $a \circ e : K \rightarrow X$  é um equalizador de  $f$  e  $g$  em  $\mathcal{E}/X$ .*

A prova segue diretamente do fato de que  $e$  é um equalizador de  $f$  e  $g$  em  $\mathcal{E}$ . Assim, *equalizadores em  $\mathcal{E}/X$  nada mais são do que equalizadores em  $\mathcal{E}$ .*

Segue-se de (I), (II) e (III) que:  *$\mathcal{E}/X$  é uma categoria finitamente completa.*

Para a cocompletude finita de  $\mathcal{E}/X$ , note primeiramente que se  $0$  é inicial em  $\mathcal{E}$ , o único morfismo  $0 \rightarrow X$  é inicial em  $\mathcal{E}/X$ .



De fato, dado um  $\mathcal{E}/X$ -morfismo  $A \xrightarrow{a} X$ , o morfismo  $0 \xrightarrow{0_A} A$  é o único morfismo que faz o diagrama

$$\begin{array}{ccc} 0 & \xrightarrow{0_A} & A \\ & \searrow & \swarrow a \\ & 0_X & X \end{array}$$

comutar  $a \circ 0_A = 0_X$ , pois  $0$  é inicial em  $\mathcal{E}$ . Assim, objetos iniciais em  $\mathcal{E}/X$  são  $\mathcal{E}$ -morfismos cujo domínio é inicial em  $\mathcal{E}$ .

Assim, temos:

(IV)  $\mathcal{E}/X$  tem objetos iniciais.

Para coprodutos e coequalizadores temos as construções análogas aos seus duais, de modo que:

(V) Um coproduto em  $\mathcal{E}/X$  é um pushout em  $\mathcal{E}$ .

(VI) Um coequalizador em  $\mathcal{E}/X$  é um coequalizador em  $\mathcal{E}$ .

Assim, temos também que:  $\mathcal{E}/X$  é uma categoria finitamente cocompleta.

(VII) Para ver que  $\mathcal{E}/X$  tem classificador de subobjetos, note primeiramente que  $1_X$  é um objeto terminal em  $\mathcal{E}/X$ . Como a identidade  $1_X$  é um monomorfismo, aplicando o  $\Omega$ -Axioma, que vale em  $\mathcal{E}$ , temos o pullback dado pelo diagrama:

$$\begin{array}{ccc} X & \xrightarrow{1_X} & X \\ |_X \downarrow & & \downarrow \chi_{1_X} \\ 1 & \xrightarrow{\top} & \Omega \end{array}$$

de forma que  $\chi_{1_X} = \top_X := \top \circ |_X : X \rightarrow \Omega$ .

Defina o  $\mathcal{E}$ -morfismo  $\top/X := \langle \top_X, 1_X \rangle : X \rightarrow \Omega \times X$ . dado pelo diagrama produto:

$$\begin{array}{ccccc} & & X & & \\ & \swarrow & \vdots & \searrow & \\ & \top_X & \langle \top_X, 1_X \rangle & 1_X & \\ & \swarrow & \downarrow & \searrow & \\ \Omega & \xleftarrow{\pi_1} & \Omega \times X & \xrightarrow{\pi_2} & X \end{array}$$

Seja  $\Omega_0 = \pi_2 : \Omega \times X \rightarrow X$ , a segunda projeção do produto.

Dados dois  $\mathcal{E}/X$ -objetos  $A \xrightarrow{a} X$  e  $B \xrightarrow{b} X$ , um monomorfismo  $\gamma : a \rightarrow b$  em  $\mathcal{E}/X$  é simplesmente um monomorfismo  $\gamma : A \rightarrow B$  em  $\mathcal{E}$  tal que o seguinte diagrama comuta:

$$\begin{array}{ccc} A & \xrightarrow{\gamma} & B \\ & \searrow a & \swarrow b \\ & & X \end{array}$$

Assim,  $\gamma$  é um subobjeto de  $B$  em  $\mathcal{E}$ .

Aplicando novamente o  $\Omega$ -Axioma em  $\gamma$  temos o seguinte pullback em  $\mathcal{E}$ :

$$\begin{array}{ccc} A & \xrightarrow{\gamma} & B \\ \downarrow & & \downarrow \chi_\gamma \\ 1 & \xrightarrow{\top} & \Omega \end{array}$$

Para transpor a construção para  $\mathcal{E}/X$ , para cada monomorfismo  $\gamma : A \rightarrow B$  em  $\mathcal{E}$  defina o  $\mathcal{E}/X$ -morfismo  $\gamma^* = \langle \chi_\gamma, b \rangle : B \rightarrow \Omega \times X$  dado pelo seguinte produto:

$$\begin{array}{ccccc} & & B & & \\ & \swarrow & \vdots & \searrow & \\ & \chi_\gamma & \langle \chi_\gamma, b \rangle & b & \\ \Omega & \xleftarrow{\pi_1} & \Omega \times X & \xrightarrow{\pi_2} & X \end{array}$$

Assim,  $\gamma^* = \langle \chi_\gamma, b \rangle$  é um morfismo em  $\mathcal{E}/X$  da forma  $\gamma^* : b \rightarrow \Omega_0 (= \pi_2)$ . Mas, então, o diagrama

$$\begin{array}{ccc} a & \xrightarrow{\gamma} & b \\ \downarrow & & \downarrow \gamma^* \\ 1_X & \xrightarrow{\top/X} & \Omega_0 \end{array}$$

é um pullback em  $\mathcal{E}/X$  e a propriedade universal provém da universalidade do  $\Omega$ -Axioma em  $\mathcal{E}$ .

Assim,  $\top/X : 1_X \rightarrow \Omega_0$  é um classificador de subobjetos para  $\mathcal{E}/X$ . Segue-se que:  $\mathcal{E}/X$  tem classificador de subobjetos.

(VIII) Resta verificar o 0-Axioma.

Lembremos que em  $\mathcal{E}/X$ , um objeto inicial é um morfismo  $0 \xrightarrow{0_X} X$ , em que  $0$  é inicial em  $\mathcal{E}$ .

Considere um  $\mathcal{E}/X$ -morfismo  $a \xrightarrow{f} 0_X$  cujo domínio é  $A \xrightarrow{a} X$ . Isto significa que o diagrama

$$\begin{array}{ccc} A & \xrightarrow{f} & 0 \\ & \searrow a & \swarrow 0_X \\ & & X \end{array}$$

comuta. Como o 0-Axioma vale para  $\mathcal{E}$ ,  $f$  é iso. Logo,  $f$  é iso em  $\mathcal{E}/X$ . Assim, temos que: *O 0-Axioma vale para  $\mathcal{E}/X$ .*

Assim, temos que  $\mathcal{E}/X$  é uma categoria com morfismos verdade. □

Para finalizar, observemos que, em teoria de topos, os axiomas (PT2) e (PT3) dos prototopos são conseqüências do teorema fundamental para esta teoria. Fica aberta a questão se esses axiomas são também válidos em categorias com morfismos verdade em virtude do teorema acima. Se esse for o caso as noções de prototopos e categorias com morfismos verdade seriam equivalentes.

## Referências Bibliográficas

- [1] Awodey, S. (2010). *Category theory*. Oxford University Press.
- [2] Bell, J. L. (2008). *Toposes and local set theories: an introduction*. Courier Corporation.
- [3] Boileau, A. e Joyal, A. (1981). La logique des topos. *The Journal of Symbolic Logic*, vol 46(1), pp 6-16.
- [4] de Souza, E. G. (2018) Sobre a noção categorial de prototopos. *Revista de Filosofia Moderna e Contemporânea*. Vol 6. PP 105-114.
- [5] Freyd, P. (1972). Aspects of topoi. *Bulletin of the Australian Mathematical Society*, vol 7, pp 1-76.
- [6] Goldblatt, R. (2006). *Topoi: the categorical analysis of logic*. Dover Publications, Inc.
- [7] Herrlich, H. e Strecker, G. E. (1973). *Category theory*. Allyn and Bacon Inc.
- [8] Kock, A. e Wraith, G. C. (1971). *Elementary toposes*. Aarhus Lecture Note Series, 30.
- [9] MacLane, S. (1971). *Categories for the working mathematician*. Springer-Verlag.
- [10] MacLane, S. e Moerdijk, I. (2012). *Sheaves in geometry and logic: A first introduction to topos theory*. Springer-Verlag.
- [11] MacLarty, C. (1995). *Elementary categories, elementary toposes*. Clarendon Press.

PARTE 2

FUNDAMENTOS DA FÍSICA E METAFÍSICA

# O GATO DE SCHRÖDINGER *não* ESTÁ VIVO E MORTO ANTES DA MEDIÇÃO: SOBRE A INTERPRETAÇÃO DOS RESULTADOS QUÂNTICOS

Décio Krause<sup>1,2</sup>

## Abstract

Muitos textos de divulgação que falam da mecânica quântica fazem referência ao "paradoxo" do gato de Schrödinger, comentando que, antes da medida do estado do sistema, o gato encontra-se vivo e morto ao mesmo tempo. Outros dizeres referem-se a que uma partícula pode estar em dois (ou mais) lugares ao mesmo tempo, que ela passou por duas fendas simultaneamente, ou ainda que percorreu dois caminhos de uma só vez, e que "interferiu consigo mesma". Na opinião deste autor, tratam-se de exageros conceituais, que podem conduzir o leitor não especializado ao erro e a imaginar situações que nada têm a ver com a mecânica quântica propriamente dita, entendida do modo como o fazemos neste artigo. Este texto de caráter geral, sem mencionar essas fontes, tem a intenção de esclarecer ou ao menos de questionar esses e outros pontos polêmicos na fala comum sobre os resultados da mecânica quântica.

## 1. Introdução

“Cats seem to play an important role in physics.”

B. Kanitscheider (1992)

Quando nos deparamos com situações inusitadas, com as quais nunca tivemos contato anterior, tendemos a procurar explicações dentro dos contextos que nos são familiares. Assim, se vemos uma imagem difusa de um animal, na tentativa de descobrir do que se trata, procuramos em nossos arquivos aqueles animais conhecidos que em nossa opinião mais se aproximam daquilo que vimos difusamente. Muitas vezes, cometemos o erro de achar que o tal animal é de algum tipo, e podemos estar equivocados. Isso parece bastante evidente para qualquer um que se ocupe a pensar sobre o assunto, e acontece também com os resultados que obtemos em nossas teorias científicas, em especial nas teorias da física. A física quântica é um dos campos do saber que mais nos apresentam resultados que nos parecem totalmente estranhos, e as explicações que tentamos dar, tomando por base o que (supostamente) conhecemos, a física clássica, via de regra não são satisfatórias. Aliás, foi precisamente por esse motivo que a física quântica surgiu, inicialmente com a hipótese de quantização de Planck para explicar os resultados observados na radiação de um corpo negro, pois a teoria vigente não explicava os fenômenos observados.<sup>3</sup>

---

<sup>1</sup> Departamento de Filosofia. Universidade Federal de Santa Catarina. Pesquisador do CNPq. Email: [deciokrause@gmail.com](mailto:deciokrause@gmail.com).

<sup>2</sup> Agradeço a Alexandre Costa-Leite pelo convite para escrever este artigo em comemoração aos cinco anos do *Seminário Permanente Lógica no Avião*, realizados na Universidade de Brasília.

<sup>3</sup> Há muitos textos excelentes de história da física quântica, como Baggott (2013), Jammer (1974), Kumar (2009).

Para discutir esse assunto ao menos de forma introdutória e para colocar nosso ponto de vista de que nos enganamos também na interpretação dos resultados em mecânica quântica, e igualmente visando alertar o leitor para equívocos, é conveniente distinguir entre uma *formulação* de uma teoria científica (aqui ficaremos restritos à física quântica)<sup>4</sup> e uma *interpretação* dessa formulação, a qual nos fornece, por assim dizer, uma *visão de mundo*. Há muitos modos de se formular uma MQ (em Styer et al. 2002, *nove* alternativas são apresentadas) que são, ao que tudo indica, equivalentes empiricamente (ainda que isso seja discutível), e a cada uma delas pode-se associar (pelo menos em princípio) várias interpretações.

A formulação padrão, usualmente considerada pelos filósofos, faz uso da *teoria dos espaços de Hilbert*, tendo sido originada com von Neumann na década de 1930. Quanto às interpretações, a considerada "padrão" é a dita "de Copenhague", devida principalmente a Bohr, Heisenberg, Born, Jordan e Pauli, ainda que haja variações mesmo entre eles. Nesse formalismo, que é o que consideremos exceto se dito explicitamente o contrário, os *estados* dos sistemas físicos são considerados e há "observáveis" que podem ser *medidos*; quando um observável é medido, considerando-se um certo estado (descrito por um vetor ou por uma *função de onda*), um valor que representa uma *probabilidade* é obtido. Assim, o que se vê é um tratamento matemático que tem como conceitos primitivos as noções de *estado* e *observável*; os sistemas físicos propriamente ditos, como partículas elementares, átomos e moléculas, aparecem apenas implicitamente, a sua natureza sendo relegada às interpretações.

O formalismo, ou os formalismos (para empregar um termo caro aos físicos) é bem entendido. O problema maior aparece quando se tenta entender o que está acontecendo. Aí é que aparecem as discrepâncias para com aquilo que conhecemos, pois as entidades tipicamente quânticas parecem se comportar de um modo totalmente inusitado e estranho à física clássica e à nossa intuição. O assunto se complica mais ainda quando explicações destinadas ao grande público são postas, quando (na opinião deste autor) muitos erros são veiculados. Analisaremos abaixo alguns deles, reduzindo o formalismo matemático a um mínimo.<sup>5</sup>

## 2. O formalismo

Segundo o formalismo padrão via espaços de Hilbert, se um sistema físico tem a possibilidade de estar em dois estados, descritos por vetores  $|\psi_1\rangle$  e  $|\psi_2\rangle$ , postula-se que qualquer combinação linear desses "estados" é também um estado possível do sistema, digamos

$$|\psi\rangle = a|\psi_1\rangle + b|\psi_2\rangle, \quad (1)$$

com  $a$  e  $b$  números complexos tais que  $|a|^2 + |b|^2 = 1$  (ver mais abaixo). Um tal estado é o que se denomina de *superposição* dos estados (descritos por)  $|\psi_1\rangle$  e  $|\psi_2\rangle$  (os abusos de linguagem serão cometidos doravante, como a identificação do vetor com o estado do sistema).

<sup>4</sup> Por "física quântica" entendemos um grupo de teorias que prioritariamente lidam com o nosso "mundo microscópico", ainda que, no limite, a física clássica, que lida com os objetos do nosso cotidiano, se reduza à quântica. Dentre essas teorias, podemos discernir entre a mecânica quântica não relativista (MQ), a mecânica quântica relativista, que incorpora a relatividade restrita (MQR).

<sup>5</sup> Um excelente livro para o público geral, que mostra as esquisitices do mundo quântico é Gilmore (1998). Textos mais bem elaborados e escritos em português, constam dos dois volumes de Pessoa Jr. (2003) e (2006) e de Chibeni (1997).

Um caso particular de uma superposição é quando temos o *emaranhamento* de dois sistemas; se temos dois estados possíveis,  $A$  e  $B$  e dois sistemas, descritos por  $|\psi_1\rangle$  e  $|\psi_2\rangle$ , um sistema emaranhado é um vetor da forma (o sinal "+" é usado para bósons e o "-" para férmions):<sup>6</sup>

$$|\psi_{12}\rangle = \frac{1}{\sqrt{2}} \left( |\psi_1\rangle^A \otimes |\psi_2\rangle^B \pm |\psi_2\rangle^A \otimes |\psi_1\rangle^B \right), \quad (2)$$

onde o coeficiente  $1/\sqrt{2}$  indica que, quando medirmos alguma coisa sobre o sistema conjunto, obtemos  $|a|^2 = |b|^2 = 1/2$  de encontrarmos o sistema em um dos estados da superposição acima, ou seja,  $|\psi_1\rangle^A \otimes |\psi_2\rangle^B$  ou  $|\psi_2\rangle^B \otimes |\psi_1\rangle^A$ , os quais indicam respectivamente o sistema é tal que o primeiro componente está no estado  $A$  e o segundo componente está no estado  $B$  ou que o primeiro está no estado  $B$  e o segundo em  $A$ . O produto " $\otimes$ " dos vetores indica o *produto tensorial*, uma operação que se pode definir entre espaços e entre vetores. A questão é saber o que indica o estado  $|\psi_{12}\rangle$  *antes* do sistema ser medido.<sup>7</sup> A teoria diz que, depois da medição de um observável, como spin ou posição, do sistema no estado  $|\psi_{12}\rangle$ , o sistema entra *imediatamente* em um dos estados componentes. Há vários modos de se denominar essa alteração, a principal chamada de *colapso da função de onda* (ou do vetor de estado).

Suponha que, depois do colapso, tenhamos obtido o estado  $|\psi_1\rangle^A \otimes |\psi_2\rangle^B$ . Aparentemente, isso nos indicaria que o *primeiro sistema* está no estado  $A$  e que o *segundo sistema* está no estado  $B$ , havendo portanto uma distinção entre eles. Mas não é assim. Devido a problemas com a *identidade* dos sistemas quânticos, algo do qual falaremos brevemente à frente, não há como distinguir um sistema do outro se eles forem do mesmo tipo (digamos, ambos sendo elétrons). Isso é devido a um postulado adicional chamado de Postulado da Indistinguibilidade (French & Krause 2006, *passim*), que diz exatamente isso: a medida de um observável para um sistema (podendo comportar vários subsistemas) em um certo estado é o mesmo (dá a mesma probabilidade) antes e depois de uma permutação dos mesmos. Ou seja, *permutações (de sistemas de mesmo tipo) não são observadas*. Que isso *deve* ser assim tinha sido apontado já por John Dalton em 1808 na origem da química moderna (veja Krause & Arenhart 2018); não pode haver diferenças entre átomos da mesma substância, ou entre seus componentes, desde que sejam de mesmo tipo (esses foram descobertos depois, como elétrons, prótons e outros.)

Mas, antes da medição, os sistemas quânticos estão em um estado que é uma *superposição* de outros estados.<sup>8</sup> Isso tudo é matematicamente claro, e o cálculo das probabilidades é algo que também se sabe fazer. A matemática funciona muito bem.<sup>9</sup> O problema surge quanto se tenta *interpretar* a superposição. Muitos dizem, erroneamente na opinião deste autor, que o sistema *está em todos os estados superpostos simultaneamente*. Isso será melhor explicado abaixo. Essa suposição é calcada em conceitos advindos da física clássica, e não temos (ainda) como exprimir em palavras o que o formalismo matemático

<sup>6</sup> Bósons e férmions são as duas grandes classes de partículas. Apesar do formalismo matemático ser compatível com a existência de outros tipos de partícula, as que se conhecem caem em uma dessas duas categorias; veja French & Krause (2006), §3.8.

<sup>7</sup> O vetor (2) pode ser abreviado por  $|\psi_{12}\rangle = |\psi_1^A \psi_2^B\rangle \pm |\psi_2^A \psi_1^B\rangle$ , omitindo-se inclusive o coeficiente  $\frac{1}{\sqrt{2}}$ .

<sup>8</sup> Insistindo, matematicamente uma superposição de estados nada mais é do que uma combinação linear de vetores que representam estados possíveis. Uma combinação linear de vetores é um outro vetor que equivale a uma soma desses vetores, cada um deles multiplicado por uma constante.

<sup>9</sup> Há no entanto outro problema relacionado ao conceito de probabilidade. Mas essa é outra história.

está contando tão claramente. O erro de se tentar usar a linguagem usual, calcada na física clássica, resulta do fato de que, como dizia Paul Dirac, essa situação não pode ser descrita por essa linguagem (Dirac 1967, Cap.1), com a qual estamos mais acostumados. Qualquer tentativa de explicar o que está acontecendo utilizando conceitos comuns conduz a absurdos como afirmar que, previamente à medição do estado do sistema, *ambos os estados estão presentes na descrição do sistema conjunto*, ou seja, que *ele está em ambos os estados ao mesmo tempo*. Vamos explicar isso a seguir, utilizando o célebre exemplo do gato de Schrödinger, amplamente discutido na literatura.

Fica aqui a mensagem deste artigo. É patente que pode-se associar ao formalismo matemático da mecânica quântica uma plêiade de interpretações não equivalentes e não há, a princípio, qualquer argumento que faça uma delas prevalecer sobre as demais. No entanto, acreditamos que se não podemos agir "positivamente", podemos fazê-lo "negativamente": podemos assegurar que pelo menos um modo de se interpretar os resultados quânticos é equivocado, a saber, aquele que supõe que, previamente à medição, o sistema se encontra simultaneamente em todos os estados possíveis. Veremos isso a seguir com mais detalhes.

### 3. O gato

Talvez o melhor exemplo que se possa usar para atestar os equívocos mencionados é o caso do célebre *gato de Schrödinger*. Em 1935, Erwin Schrödinger criou um experimento mental para enfatizar os aspectos contra-intuitivos das superposições e da dominante (à época) "interpretação de Copenhague". Ele imaginou o experimento envolvendo um objeto macroscópico, um gato, justamente para enfatizar o seu ponto. Porém, saliente-se, não se observam objetos macroscópicos em superposição, mas isso ocorre com objetos à escala quântica.<sup>10</sup>

Abreviadamente, a questão é a seguinte. Imagine um gato preso em um quarto hermeticamente fechado no qual há um frasco de veneno e uma amostra de material radioativo que tem 50% de possibilidade de emitir uma partícula (decair) no decorrer de uma hora. Se a partícula for emitida, um mecanismo é ativado de modo que o frasco de veneno é quebrado, o veneno é liberado e o gato morre. Se nenhuma partícula for emitida, o frasco continua intacto e o gato permanece vivo. Para saber como está o gato, ou seja, qual é o seu estado depois de uma hora, o cientista abre a porta do quarto e observa o gato. Isso corresponde a uma *medição*, a medir o estado do gato exatamente no momento em que o cientista abre a porta. O sistema é preparado de forma que há probabilidade de 1/2 de o cientista encontrar o gato vivo e 1/2 de encontrá-lo morto (essa possibilidade é correspondida empiricamente). O estado do gato previamente à abertura da porta é descrito por um vetor semelhante a (2), expressando uma soma (ou subtração) de um vetor (um produto tensorial) que representa a situação de que o material não decaiu e o gato está vivo com outro vetor representando a situação de que o material decaiu e o gato está morto.

Segundo a interpretação usual (e equivocada), antes da medição o gato está vivo e morto ao mesmo tempo, partilhando, a uma só vez, de ambos os estados possíveis.

<sup>10</sup> Quando os sistemas têm *muitos* elementos quânticos, como um gato, que é composto de trilhões de átomos e moléculas, a ação do meio ambiente ocasiona um fenômeno ao qual os físicos dão o nome de *decoerência*, ou seja, o sistema "perde" a sua "coerência" quântica, comportando-se como um "sistema clássico". Veja Schlosshauer (2004).



No caso, temos uma situação particular de uma superposição, denominada de *emaranhamento*, que Schrödinger chegou a dizer (em 1935) ser a característica distintiva da mecânica quântica, sem similar na física clássica (Schrödinger 1980).

Experimentos que simulam a situação do gato de Schrödinger podem ser realizados ao nível quântico, como tem sido comum nos dias de hoje, colocando-se sistemas quânticos em estados emaranhados. Essa é, na verdade, uma das maiores revoluções proporcionadas pela MQ, possibilitando, dentre outras coisas, a computação quântica e as discussões sobre a informação quântica.

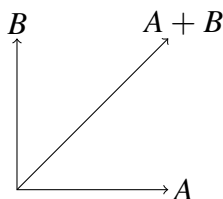
**3.1 Ontologia.** Mas, antes da medição, qual a situação do gato? Antes de discutir isso, é preciso conceder que podemos falar do gato, ou seja, que o nosso uso da mecânica quântica não é meramente instrumentalista, como advogam alguns físicos a começar por Niels Bohr, mas que há algum tipo de ontologia associado à teoria.<sup>11</sup> É sabido que Bohr via com suspeição a possibilidade de uma realidade independente dos fenômenos quânticos que se assemelhe àquela descrita por uma ontologia aos moldes clássicos, de objetos tais como os descritos pela física clássica.<sup>12</sup> Assim, temos que adotar uma postura diferente e admitir que, mesmo antes da medição, há algo do que podemos falar. A questão ontológica é especificar em que consiste esse algo, e a resposta não é única. Mas esse é outro assunto.<sup>13</sup>

Para os instrumentalistas, a mecânica quântica não nos permite falar de gatos ou de partículas; ela é meramente uma teoria matemática que nos permite computar probabilidades, como a de que o estado do sistema sendo uma superposição "gato+material radioativo", após a medição, resulte algo como

$$|\text{gato vivo}\rangle \otimes |\text{não houve decaimento}\rangle,$$

o que atestaria que não houve decaimento e o gato continua vivo. Só isso, nada mais.

**3.2 Superposições.** Vamos ser mais precisos. Antes da medição (ou seja, do cientista abrir a porta), o estado do sistema conjunto gato mais material radioativo é descrito por um vetor da forma  $A + B$ , onde  $A$  é o vetor que indica o estado em que o material radioativo não decaiu e o gato está vivo, enquanto que  $B$  é o vetor que indica a situação em que o material emitiu radiação e o gato está morto.



<sup>11</sup> Na verdade, como argumentaremos abaixo, há várias ontologias possíveis de serem associadas a uma teoria.

<sup>12</sup> Alguns autores creditam a Bohr uma visão *empirista* da MQ. Outros, como Folse Jr (Folse Jr 1986) e o presente autor, preferem vê-lo não como rejeitando um realismo (exceto aquele no sentido de Einstein — ver a seção 3.3, mas como possibilitando a admissão de uma realidade de entidades *distintas* daquela apregoada pela física clássica. Ou seja, Bohr poderia ser um realista de entidades (no sentido de Hacking 1983), mas de *entidades de "outro" tipo*, aquelas "tipicamente quânticas".

<sup>13</sup> Com efeito, a natureza dessas entidades vai depender da interpretação que se adote. À formulação da MQ devida a David Bohm, uma "interpretação clássica" pode ser associada, no sentido de que as entidades quânticas não diferem essencialmente daquelas descritas pela física clássica; para uma distinção entre *ontologia clássica* e *ontologia quântica*, ver Krause et al. (2012).

Fig. 1 – A soma de dois vetores ortogonais não nulos produz um terceiro vetor diferente de ambos.

O erro frequente na interpretação desse resultado é não considerar que a soma  $A + B$  é um terceiro vetor, que indica uma superposição dos estados  $A$  e  $B$ , e que é distinto de ambos (supostos não nulos). Aqui acaba a matemática, entrando a interpretação. No caso do gato, o que  $A + B$  quer dizer? Lembre mais uma vez de Dirac: isso não pode ser explicado em termos clássicos, ou seja, não há analogia possível com algo que conhecemos em nosso dia a dia. Se podemos falar no gato (deixando de lado o instrumentalismo), o seu estado é descrito pela superposição de  $A$  e  $B$ , e não por esses vetores isoladamente, e muito menos pela sua "conjunção", que indicaria que ele estaria simultaneamente nos dois estados. É errado portanto dizer que o gato está vivo (o que exigiria que o seu estado fosse  $A$ ) ou que está morto (o que exigiria que seu estado fosse  $B$ ), e muito menos "ao mesmo tempo", o que exigiria que considerássemos a sua *conjunção*, e não a sua soma! O estado do gato é dado pelo vetor  $A + B$ , portanto nem por  $A$  e nem por  $B$ , mas pelo *terceiro* vetor, que não é uma conjunção lógica, mas uma soma vetorial.<sup>14</sup>

O estado do gato, previamente à medição, portanto, é descrito pelo vetor  $A + B$ . Depois da medição, esse vetor soma *colapsa* (segundo a interpretação dominante) por um motivo que ninguém conhece, em um dos estados componentes,<sup>15</sup> mas antes disso temos somente a probabilidade de que o estado resultante será um ou outro, *nunca ambos*. Fiquemos portanto com a seguinte mensagem: previamente à medição, o gato não está vivo e morto: *não há contradição nesse caso*.

Da mesma forma, muitos textos se referem à possibilidade de uma partícula (ou melhor, um sistema quântico) estar em dois ou mais lugares ao mesmo tempo (mais na seção 5). Outro equívoco: esses sistemas não têm posição em sentido usual, mas o que poderíamos chamar de *estados-posição*. Antes da medição, o sistema está em uma superposição de estados que indicam, por meio de uma probabilidade, onde ele (reparem no nosso discurso, tendo que nos referi a "ele" como se fosse um objeto descrito pela física clássica) poderá aparecer depois que sua posição for medida, por exemplo, em um determinado lugar em um anteparo que recebe as partículas. Portanto, guarde isso também: sistemas quânticos *não podem estar em mais de uma posição a um só tempo*. A linguagem não é precisa, mas a matemática é bem clara, expressa pela superposição de todas as posições possíveis e suas probabilidades, e não de uma conjunção de quaisquer coisas que indiquem posições.

**3.3 O realismo de Einstein.** Fizemos referência acima ao fato de que Bohr contestava algumas posições de Einstein. Este acreditava na existência de uma realidade externa calcada nos conceitos de *separabilidade* e *localidade* (Howard 2017). Para ele, sistemas

<sup>14</sup> Dizendo isso, somos contrários à interpretação *paraconsistente* da MQ dada por da Costa e de Ronde (2013), que admitem a possibilidade de haver uma contradição no caso do gato previamente à medição. Discutimos a posição desses autores em Arenhart & Krause (2016). Como dissemos na Introdução, não obstante em geral não poderemos optar por uma interpretação exceto por motivos nada objetivos, *essa* é uma interpretação que em nossa opinião não pode ser aceita. Se a MQ admite alguma contradição, certamente ela não está em situações como as descritas aqui e abaixo. No entanto, veja a seção 4.

<sup>15</sup> Na interpretação dos *muitos mundos* (Wallace 2012), quando há a medição, o universo se divide em vários, não havendo o colapso; no caso do gato, em dois, um no qual o gato está vivo e outro no caso em que o gato está morto. Mas mesmo assim não há qualquer contradição, pois o gato nunca está vivo e morto em um mesmo mundo.

localizados em posições distintas do espaço, a uma distância  $d > 0$  um do outro, deveriam ser considerados como sistemas físicos distintos e qualquer ação em um deles só poderia influenciar o outro se algum tipo de informação fosse transmitida, como um fóton, o que (pela relatividade restrita) não pode acontecer *imediatamente*, devido ao fato da velocidade da luz ser finita. A mecânica quântica, por outro lado, considera que mesmo sistemas separados podem ser considerados como um só sistema, uma vez que tenham interagido no passado. Em tais sistemas *emaranhados*, a ação sobre um deles permite que se saiba fatos do outro imediatamente (isso, no entanto, não viola a relatividade restrita).<sup>16</sup>

O que se chama de "realismo de Einstein", no entanto, é a crença de que todos os observáveis físicos que um sistema têm valores bem determinados em qualquer instante de tempo como na física clássica, ainda que não os conheçamos objetivamente. Que isso não é assim na MQ é fato bem conhecido, em especial devido a um teorema demonstrado em 1967 por Simon Kochen e Ernst Specker (antecipado por John Bell em 1966), mas isso não será comentado aqui. Além disso, para Einstein, a física deve ser *determinista*: dada a situação de um sistema em um dado instante, as equações devem nos possibilitar conhecer a situação do sistema em qualquer instante de tempo posterior (ou anterior). A formulação de Copenhague, pelo contrário, não é determinista nesse sentido porque não consegue prever com 100% de segurança o que acontece com o estado do sistema depois de uma medição (depois do "colapso"), oferecendo apenas probabilidades.<sup>17</sup>

#### 4. A negação quântica

O problema parece depender do modo como se entendem "morto" e "vivo" (a negação de "morto"), no caso do gato. Usualmente, entende-se por *contradição* a conjunção de duas proposições (sentenças ou fórmulas), uma das quais sendo a negação da outra. Se usarmos " $\neg$ " para representar a negação, " $\wedge$ " para a conjunção e " $S$ " para denotar uma sentença, uma contradição é uma sentença da forma " $S \wedge \neg S$ ". As sentenças " $S$ " e " $\neg S$ " dizem-se *contraditórias*. Na linguagem da maioria das lógicas, como no caso da lógica clássica, se temos ambas  $S$  e  $\neg S$ , podemos formar a sua conjunção e obter  $S \wedge \neg S$ .<sup>18</sup> Por outro lado, se temos uma contradição, podemos derivar ambas as sentenças. Repare que não faz sentido falar "ao mesmo tempo" a não ser como abuso de linguagem (ou que tenhamos algum tipo de lógica temporal). A lógica clássica, assim como a maioria dos sistemas lógicos usuais, não envolve a noção de tempo.

Assim, insistamos nisso, se temos uma contradição, temos que ter as duas situações descritas pelas sentenças  $S$  e por sua negação,  $\neg S$ . Dito informalmente, ambas têm que ser "verdadeiras" (deixemos o sentido do vocábulo "verdade" implícito, interpretando-o informalmente). Portanto, no caso do gato de Schrödinger, se dizemos que (previamente à medição) o gato está vivo e morto, temos que dispor dos dois estados *vivo* e *morto* como

<sup>16</sup> Não é nosso objetivo discutir esse importante ponto aqui. Veja Kumar (2009) sobre o célebre debate entre Bohr e Einstein sobre o tema. O importante é saber que é errado dizer que a medição em um sistema *influencia* o outro; nenhuma "influência" existe pelo motivo explicado. O que acontece é algo tipicamente quântico; a medição em um sistema permite saber o que acontece com o outro, *sem* que qualquer mensagem tenha sido emitida. Este é reconhecidamente o maior dos mistérios quânticos, mas é um fato mais do que estabelecido empiricamente nos dias de hoje.

<sup>17</sup> Um excelente texto geral sobre a filosofia de Einstein é Howard (2017); ver também Fine (1986).

<sup>18</sup> Doravante não faremos mais a distinção entre uso e menção.

estados *atuais* do gato, o que não ocorre. O seu estado, como vimos, é descrito por uma superposição, uma soma vetorial, e não por uma conjunção lógica.

Que isso é assim já foi mencionado por importantes autores, como Beltrametti & Cassinelli (1981, pp.220-1) e por Bas van Fraassen (1974, p.229) no contexto das lógicas quânticas (mas veja também Aerts et al. 2000). Tanto os dois primeiros como o terceiro mencionam que a negação em mecânica quântica deve ser vista não como *contraditoriedade* (uma sendo verdadeira se e somente se a outra for falsa), mas como o que van Fraassen chamou de *choice negation*.

**4.1 Negações clássica e quântica.** Para van Fraassen (assim como para Beltrametti e Cassinelli), a *choice negation* permite que sejam discernidos os conceitos de falsidade e de não-verdade, que são tomados como equivalentes com a negação usual, que van Fraassen denomina de *exclusion negation* (uma sentença exclui a outra). Na física quântica, por outro lado, podem haver situações em que tanto uma quanto a outra possam ser falsas, mas não possam ser ambas verdadeiras, como por exemplo quando o estado é uma superposição de estados. Explicando melhor: no caso em que o estado do gato é uma superposição de estados ou que a posição de uma partícula é uma superposição de posições possíveis (fiquemos com o caso do gato), ambas "vivo" e "morto" são falsas, pois o seu estado não é qualquer desses, mas descrito pela soma vetorial, por um vetor como (2), como vimos acima (o mesmo ocorre com as posições, obviamente).

Podemos definir semanticamente uma *choice negation* ou, como preferimos denominar, uma *negação quântica* do seguinte modo. Suponha que dispomos de uma linguagem adequada, digamos de primeira ordem, contendo um símbolo de negação  $\neg_q$ . Seja  $V$  o conjunto dos valores de verdade, sendo  $V = D \cup N$ , com  $D \cap N = \emptyset$ . O conjunto  $D$  é o dos valores distinguidos, e  $N$  o dos valores não distinguidos; se quisermos, podemos tomar  $V = \{0, 1\}$ , com  $D = \{1\}$  e  $N = \{0\}$ . Se  $\alpha$  é uma fórmula e  $\nu$  é uma valoração, ou seja, uma função  $\nu : \mathcal{F} \mapsto V$ , onde  $\mathcal{F}$  é o conjunto das fórmulas, pomos:

$$\nu(\neg_q \alpha) \in D \text{ implica } \nu(\alpha) \in N. \quad (3)$$

Como a recíproca não vale necessariamente, fica claro que tanto  $\alpha$  como  $\neg_q \alpha$  podem assumir valores não distinguidos (serem ambos "falsas") sem que tenham que assumir também valores distinguidos — um estudo mais detalhado é feito em Krause & Gracher, a aparecer.

Baseados em uma interpretação advinda do Quadrado de Oposições, Arenhart e este autor deram uma interpretação para a *choice negation* em termos de contrariedade (Krause & Arenhart 2016), algo já adiantado em Priest & Routley (1989), Béziau (2006), apesar desses autores não terem se referido à mecânica quântica especificamente. Se uma proposição e sua negação são vistas como *contrárias* e não como *contraditórias* (negação exclusiva, representada nas diagonais do quadrado), a intuição da *choice negation* é obedida: ambas podem ser falsas, sem que possam ser ambas verdadeiras (no caso do gato, a terceira situação ocorre quando o estado é descrito por uma superposição de estados).

Nota-se observando o quadrado que, devido à sub-alternação,  $\neg_q \alpha$  implica  $\neg \alpha$ , como lógicos quânticos dizem ser o caso (Aerts et al. 2000); a demonstração pode ser vista em Krause & Gracher (a aparecer).<sup>19</sup>

<sup>19</sup> A sub-contrariedade é vista como uma outra negação, dita *negação paraconsistente*.

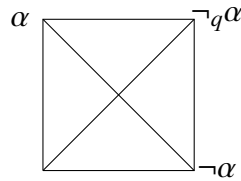


Fig. 2 – O quadrado das oposições exibindo a contrariedade. Proposições contrárias, como  $\alpha$  e  $\neg_q\alpha$  podem ser ambas falsas, mas não podem ser ambas verdadeiras. A negação "clássica" de  $\alpha$  é indicada por  $\neg\alpha$ , sua *contraditória*.

Com efeito, as sentenças  $A$  (o material radioativo não decaiu, gato vivo) e  $B$  (o material decaiu, gato morto) são contrárias, e não contraditórias. A negação de "gato vivo" não é "gato morto" e vice-versa, pois "não estar vivo (morto)" pode significar que está em um estado que é uma superposição desses dois estados, ou seja, como advogou van Fraassen, "não estar vivo" pode não equivaler a "estar morto" (a negação exclusiva de "estar vivo"). Insistindo, isso indica que ambas podem ser falsas, mas não podem ser ambas verdadeiras.

No algebrismo quântico, os vetores  $A$  e  $B$  da figura (1) são *ortonormais* (ortogonais e unitários) e a ortogonalidade é o que se chama de uma *involução*, e semanticamente não coincide com a negação da lógica clássica (Beltrametti & Cassinelli op.cit, p.221; Aerts et al. 2000). Em outros termos, se isso faz sentido, negar o estado  $A$  significa tomar qualquer estado distinto de  $A$ , mas não necessariamente  $B$ . O mesmo se pode dizer de  $B$ .

### 5. Em dois lugares ao mesmo tempo?

Voltemos ao caso da posição de uma partícula, agora com um pouco mais de detalhes. Como dissemos acima, muitos textos falam que, previamente a uma medição, uma partícula pode estar em dois ou mais lugares ao mesmo tempo, o que também é equivocado. Da mesma forma que no caso do gato, uma partícula não tem posição em sentido clássico, mas *estados-posição*, se podemos chamar assim. Aliás, a suposição usual pressupõe que uma partícula seja algo como uma pequena bolinha que estaria "por aí", no caso, em vários lugares ao mesmo tempo. A mecânica quântica não corrobora essa interpretação (como vimos insistindo, essa disciplina nada afirma sobre a natureza das entidades). Ou seja, antes do vetor que descreve sua posição colapsar por causa de uma interação (medição), o que há é um vetor que também é uma superposição de todos os estados (posições) possíveis, e os quadrados dos coeficientes dessa combinação linear indicam a probabilidade do vetor colapsar naquele estado associado ao coeficiente. Assim, se um vetor de estado é algo como (não necessariamente expressando posições)

$$E = aA + bB + cC + dD,$$

após o colapso o estado reduz-se a  $A, B, C$  ou  $D$  respectivamente com probabilidades  $a^2, b^2, c^2$  e  $d^2$ . Portanto, nada de duas ou mais posições ao mesmo tempo antes ou depois da medição.

*Note bem* – trabalhamos, como na combinação linear acima, com *amplitudes de probabilidades*, cujos quadrados fornecem as probabilidades estrito senso. Isso pode ser (como geralmente é) interpretado como indicando que alguma coisa existe previamente à medição, e que sua localização só será determinada após a medição; antes, só temos probabilidades. Esse discurso, no entanto, não é adequado pois sugere que exista algo (uma bolinha?) "escondido", o que não parece ser o caso. Voltaremos a esse ponto mais abaixo.

## 6. Os orbitais e o uso de figuras

Outro equívoco, relacionado com o que se acabou de dizer, vem do uso de imagens e de sua interpretação. Ocorre em física o mesmo que ocorre em matemática: figuras têm unicamente papel heurístico, não podendo ser utilizadas para se fazer inferências, exceto em linguagens propositadamente alteradas para isso, como em teoria de categorias e nos diagramas de Feynman.

No entanto, em muitos livros de química vemos desenhos dos orbitais como áreas borradas e usualmente se diz serem *as regiões onde há maior probabilidade de se encontrar o elétron*. Outro equívoco, porque passa a idéia de que o elétron seria uma coisinha como uma pequena bola escondida naquela fumaça (veja a Figura 3 abaixo). Não é nada disso; o elétron não tem uma imagem clássica. A região sombreada corresponde a uma representação gráfica do orbital  $1s$  de um átomo de hidrogênio, mas de modo mais preciso corresponde ao que seria a imagem da função de onda normalizada  $\psi_{100}$  que tem como números quânticos  $l = 1, m = n_l = 0$  — para detalhes, Griffiths (2011), §4.2. O quadrado dessa função, integrada sobre um volume  $dV$  indica, como se diz, "a probabilidade de se encontrar o elétron no dado volume". Esse discurso, como vimos insistindo, é dúbio. Trata-se de um nível de energia para o átomo: nada é dito sobre o formato do elétron, muito menos que ele é ou deve ser algo "escondido" nessa região. Como diz Griffiths (op. cit., p.116), a região é uma superfície de uma densidade de probabilidade constante. É isso o que temos; qualquer inferência sobre a natureza do elétron é nossa, faz parte da interpretação.

Dizer que o elétron é a região toda também não resolve o problema, pois estaríamos identificando algo supostamente *físico* (o elétron) com uma abstração matemática (a função de onda ou eventualmente — e com maior equívoco ainda — com sua representação gráfica).<sup>20</sup>

A natureza dos objetos quânticos é algo discutível, mas aceita-se que se necessita de uma ontologia que não tem nada a ver com a clássica (em Krause et al. (2012), damos alguns detalhes do que seria uma ontologia quântica em comparação com uma ontologia clássica).

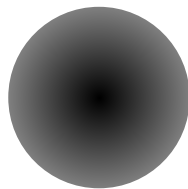


Fig. 3 – Esquema do orbital  $1s$ . No discurso comum, a área mais escura é aquela "onde é mais provável encontrar o elétron".

<sup>20</sup> A analogia com um conjunto difuso igualmente não é boa porque os elementos desses conjuntos ainda são objetos dotados de identidade (ver abaixo), havendo somente uma ignorância em sentido epistemológico sobre *onde* eles estão localizados. No caso dos elétrons, a situação é outra; não temos imagens dessas coisas, mas unicamente sua descrição matemática, dada pela *função de onda* que está concentrada em torno do núcleo mas não tem *ponto fronteira*. (Um ponto fronteira de um conjunto é um elemento do conjunto tal que qualquer círculo – para ficarmos no plano – com centro no ponto tem tanto elementos do conjunto como elementos de fora do conjunto).

## 7. As duas fendas

Outro erro conceitual, na opinião deste autor, ocorre quando se tenta "explicar" o famoso experimento das duas fendas. É comum o discurso de que, quando ambas estão abertas, a partícula (o objeto quântico, seja lá o que ele for) passa pelas duas fendas ao mesmo tempo, e que quando eles são enviados um a um, "eles interferem com eles mesmos", isso para que se possa explicar o fenômeno da interferência.<sup>21</sup> Nada disso é dito pela mecânica quântica. Novamente, antes da partícula chegar ao anteparo, quando ocorre o colapso, o que temos é um estado de superposição dos estados "passar pela fenda 1" e "passar pela fenda 2", mas isso não indica que temos de uma só vez as duas situações isoladamente. De novo, é preciso constatar que a superposição " $|\text{passar pela fenda 1}\rangle + |\text{passar pela fenda 2}\rangle$ " não é a conjunção de suas sentenças, mas uma soma vetorial. A mecânica quântica não fala nada da partícula antes de ela alcançar o anteparo. Isso é coisa da interpretação dada, que, de novo pensa a partícula como uma pequena bolinha.

Com efeito, há algumas coisas que a mecânica quântica nos conta e que vão de encontro à concepção usual que temos dos objetos físicos. Algumas delas são as seguintes, e se referem ao elétron e ao experimento das duas fendas quando padrões de interferência são observados. Nesse caso, são falsas as afirmativas seguintes: a) o elétron passa só pela fenda 1 b) o elétron passa só pela fenda 2 c) o elétron passa pelas duas fendas d) o elétron não passa por fenda alguma.

As razões são, de modo breve, as seguintes: uma vez que se verificou que o anteparo apresenta franjas de interferência (veja a Figura 4), a) e b) têm que ser falsas, pois se uma delas fosse verdadeira (o que implicaria que a outra seria falsa), o padrão observado não seria o de interferência. c) não pode ser verdadeira pois não se pode "quebrar" uma partícula em duas, exceto, de certo modo, na mecânica Bohmiana, onde cada partícula é "guiada" por uma onda, assim havendo as duas coisas acopladas (veja a seção ()). A sentença d) também não pode ser verdadeira, pois verificou-se que há registro no anteparo; logo, algo passou. O problema, como o vemos, é que não estamos lidando com uma entidade como aquelas descritas pela física clássica e que se conforma com os objetos de nossa intuição.

Do ponto de vista lógico, isso traz um problema, que discutiremos na seção 9. Os lógicos quânticos, como em Aerts et al. (2000), creem no entanto que a conjunção nas lógicas quânticas se comporta como a conjunção da lógica clássica. Aceitando essa premissa, como explicar a discrepância apresentada pelos quatro itens acima? "Classicamente", eles não poderiam ser todos violados. Nossa explicação é a de que a negação das quatro possibilidades se dá no nível metateórico, no nível da interpretação, e não na linguagem objeto. Se não houvesse discrepância entre os níveis metateórico (das interpretações) e teórico (do formalismo matemático), certamente não seria possível atribuir a um mesmo formalismo interpretações distintas, mas apenas uma interpretação "necessária".<sup>22</sup>

<sup>21</sup> Não vamos revisar o experimento aqui. Uma descrição excelente, ainda que contendo o vício para o qual estamos chamando a atenção, pode ser vista em Kumar (2009), p.267ss.

<sup>22</sup> Em French & Krause (2006), oferecemos duas interpretações para o formalismo da mecânica quântica não relativista; uma delas considera os objetos quânticos como *indivíduos*, ao par com seus correlatos na física clássica; a outra os toma como *não-indivíduos*, entidades para as quais o conceito usual de identidade não se aplica. Ademais, mostra-se que a física não nos permite determinar qual interpretação é a "correta". Como dizemos naquele livro, a metafísica fica subdeterminada pela física.

Por enquanto, como podemos entender uma situação como essa? O fato é que, usando nossos cânones "clássicos", com os quais estamos acostumados, não temos como. A situação é completamente nova, completamente quântica. O elétron não é algo que possa ser comparado a um objeto macroscópico qualquer, e seu comportamento não pode ser descrito pela física clássica, como tentamos fazer quando imaginamos que uma dentre as quatro afirmativas acima tem que ser verdadeira. Atualmente, experimentos com as duas fendas têm sido realizados com coisas "grandes" para os padrões quânticos, como com moléculas contendo 810 átomos! (Eibenberg et al. 2013) e os padrões de interferência típico de ondas têm sido observados.<sup>23</sup>

No experimento das duas fendas, o que a mecânica quântica nos mostra são duas alternativas de situações que podem ser observadas no anteparo; (1) uma situação envolvendo franjas de interferência e (2) apenas uma dentre duas bandas de pontos, conforme a figura abaixo, dependendo de qual fenda esteja fechada. Porém, nada sabemos sobre o que há entre a fonte e o anteparo.

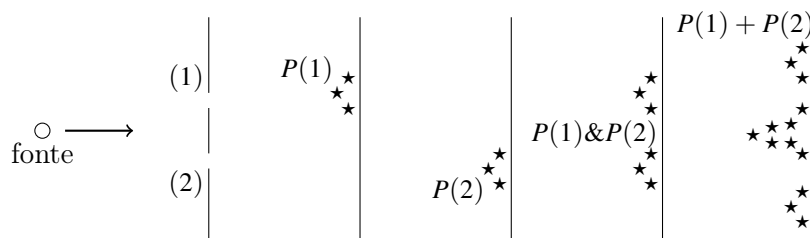


Fig. 4 – Esquema ds resultados obtidos no experimento das duas fendas. O padrão  $P(1)$ , é observado quando somente (1) está aberta; o padrão  $P(2)$  ocorre quando apenas (2) está aberta. O padrão  $P(1) \& P(2)$  seria o *esperado* para as duas fendas abertas, mas não é o que ocorre. Pelo contrário, o que ocorre é  $P(1) + P(2)$ , um padrão de interferência.<sup>24</sup>

Com efeito, a interpretação é nossa. Como dito na Introdução, sempre que enfrentamos uma situação nova à qual queremos dar uma explicação, iniciamos com o que temos às mãos. No caso, a física clássica e nosso conhecimento dos comportamentos ondulatório e de partículas, e na falta de coisa melhor, é essa a associação que fazemos. Assim, baseados em nossos conhecimentos de física clássica e do senso comum, associamos à primeira alternativa a um comportamento do tipo ondulatório, enquanto que a segunda apresenta um comportamento do tipo de partículas. No entanto, nada disso é dito pela mecânica quântica, que nos apresenta apenas os resultados matemáticos. Tudo o que a mecânica quântica nos dá são as duas alternativas, só isso. Com efeito, "ser onda" e "ser partícula" não são estados de sistemas quânticos, e nunca se vê uma superposição de "ser partícula" e "ser onda", mas a função de onda (vetor de estado) é uma superposição de "passou pela fenda 1" e "passou pela fenda 2".

Em 1970, o grande físico John Archibald Wheeler inventou uma analogia interessante. Comparou a situação com um *dragão de fumaça*, do qual conhecemos o rabo (o que acontece na fonte, por exemplo que tipo de radiação estamos emitindo, se elétrons ou

<sup>23</sup> Em Pessoa Jr. (2006), capítulo XXXI, §8, comenta-se sobre esse tipo de fenômeno com outros objetos "macroscópicos". Em Zeilinger (2005), comenta-se o experimento das duas fendas feito com objetos como fulerenos, ou *moléculas bolas de futebol*, que são moléculas compostas de 60 ou 70 átomos de carbono, algo considerado "grande" para os padrões quânticos.

<sup>24</sup> No formalismo matemático, isso é explicado quando se toma o quadrado da função de onda (que fornece a probabilidade):  $\|P(1) + P(2)\|^2 = \|P(1)\|^2 + \|P(2)\|^2 + 2\text{Re}\langle P(1)|P(2)\rangle$ , sendo  $2\text{Re}\langle P(1)|P(2)\rangle$  o *termo de interferência*, o dobro da parte real do produto interno entre  $P(1)$  e  $P(2)$ .



fótons), e a cabeça, ou seja, os resultados observados, como as situações  $P(1)$ ,  $P(2)$  ou  $P(1) + P(2)$  acima, e mais nada: tudo entre a cabeça e o rabo é fumaça, não sabemos como é! Imaginar o corpo do dragão é coisa nossa. No entanto, insistamos, não podemos imaginar "qualquer coisa". Por exemplo, supor que no interior da fumaça o sistema esteja partilhando de estados contraditórios é equivocada pelos motivos já expostos.

### 8. Mach-Zehnder

Situação semelhante ocorre com o experimento Mach-Zehnder, quando é de novo errado afirmar que a "partícula" percorre os dois caminhos ao mesmo tempo (Figura 5). A mecânica quântica nada fala do que acontece com o sistema físico antes de ele alcançar um dos receptores. Qualquer conclusão a esse respeito é coisa da interpretação. Se os dois caminhos  $A$  e  $B$  estiverem disponíveis, o que ocorre se nenhum "observador" for colocado em um deles (de modo que se possa conhecer qual caminho o objeto percorreu), a interpretação intuitiva e baseada na física clássica, que como vimos insistindo Dirac considerava equivocada, é a de que a partícula passou por ambos os caminhos, caso em que toda a radiação é detectada no receptor 1, e nenhuma no receptor 2. E isso ocorre mesmo se for emitida uma partícula de cada vez. Mas se um dos caminhos estiver interrompido ou sendo "observado", constata-se que 50% da radiação é detectada em cada receptor. Uma vez que se emite uma partícula por vez (estando ambos os caminhos sem detetores), *só se pode pensar* que a partícula percorreu ambos os caminhos de uma só vez, diz a interpretação comum.

Mas, como pode uma partícula se dividir em duas se não há "meia partícula"? Isso não faz sentido; não sabemos o que ocorre entre o rabo do dragão (a fonte) e a sua cabeça (os receptores). Não temos como explicar isso coerentemente em termos daquilo com o que estamos acostumados. Não há qualquer analogia "clássica" que possamos fazer. Trata-se de um fenômeno tipicamente quântico, sem qualquer análogo na física clássica, para a qual tentamos reduzir praticamente tudo para alcançar uma explicação. Uma descrição ainda intuitiva, porém mais detalhada, pode ser vista em Krause (2017), cap.6. Saliente-se que o comum é aceitar que a nossa ignorância acerca do corpo do dragão não é de natureza epistemológica, pois isso nos faria ter que aceitar a existência de alguma forma de variável oculta, mas (uma vez que afastemos o instrumentalismo) de natureza ontológica. Colocando em termos diretos, *o bicho é outro*.

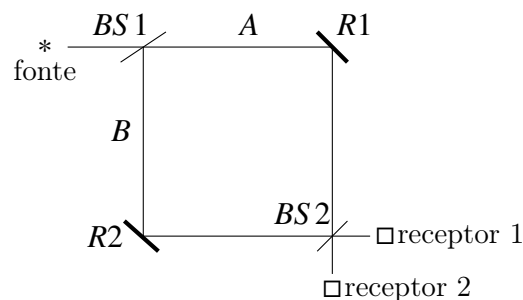


Fig. 5 – O interferômetro Mach-Zehnder. Se os dois caminhos  $A$  e  $B$  estiverem disponíveis, toda a radiação é observada no receptor 1, e nenhuma no receptor 2. Intuitivamente, isso indica que a radiação passou por ambos os caminhos. Um *beam splitter* ( $BS$ ) é um dispositivo que, em termos clássicos, divide o raio de luz ou outra radiação eletromagnética, fazendo-a percorrer os caminhos disponíveis.  $R1$  e  $R2$  indicam refletores.

## 9. A identidade

Outro conceito problemático e sobre o qual muitos equívocos são postos é o de identidade. A tendência dos filósofos e cientistas em geral é não aceitar que possa haver qualquer questionamento sobre esse conceito, e que ele é *necessário* (Bueno 2014). Claro que primeiramente necessitamos saber do que estamos falando.

Temos uma noção intuitiva de identidade quando aceitamos que todos os objetos podem ser identificados, pelo menos em princípio, como sendo *aqueles* objetos com os quais já tivemos alguma experiência anterior; o computador à minha frente é *o mesmo* que o que estava à minha frente ontem. Cada um desses objetos é único e *diferente* de todos os outros. Se necessário, fazemos uma marca em cada um deles, colocamos-lhes algum rótulo identificador, o que nos permitirá discernir mesmo entre objetos muito semelhantes, como dois comprimidos de uma caixa de remédio. Os objetos que nos cercam são *indivíduos*, são vistos como unidades, têm individualidade, podem ser individualizados, isolados. Os conceitos de identidade, individualidade e individuação são geralmente tratados como sinônimos, mas apresentam diferenças, discutidas em Arenhart, Bueno & Krause (a aparecer), porém não repetidas aqui.

Há bastante tempo, David Hume colocou dúvidas sobre essa concepção intuitiva: a tendência que temos de pensar que um objeto possa permanecer o mesmo, para o qual atribuímos uma identidade, dever-se-ia a uma conexão que fazemos entre causa e efeito (constância em nossas percepções), mas não temos qualquer garantia de que ele não sofreu mudanças, mesmo que haja semelhanças que se mantenham (Hume 1985, p.74, e *passim*). Assim, se considerarmos que a noção usual de identidade, dita *identidade leibniziana*, pressupõe que nenhum objeto pode ter *exatamente* os mesmos atributos que qualquer outro, quando sofre mudanças, como avançar na idade, o objeto já não seria mais o mesmo, idêntico ao anterior. No tocante às entidades quânticas, Schrödinger disse coisa semelhante:

"quando um objeto familiar entra novamente em nosso campo de percepção, é usualmente reconhecido como uma continuação de aparecimentos prévios, como sendo a mesma coisa. A permanência relativa de peças individuais de matéria é a característica mais importante tanto da vida diária quanto da experiência científica. Se algo familiar, digamos uma moringa de barro, desaparece de nossa sala, teremos quase a certeza de que alguém a carregou. Se após um certo tempo ela reaparece, poderemos ficar em dúvida se é a mesma moringa ou não — objetos frágeis em tais circunstâncias frequentemente não são. Podemos não ser capazes de decidir a questão, mas devemos não ter dúvida de que a questionável similaridade tem um significado indisputável — a de que há uma resposta não ambígua à nossa pergunta. Tão certa é a nossa crença na continuidade das partes não observadas das sequências!" (Schrödinger 1998).<sup>25</sup>

Com efeito, na tentativa de encontrar o motivo das coisas nos parecerem indivíduos, somos levados à procura de algum Princípio de Individuação. A literatura filosófica nos apresenta dois grandes grupos de teorias a esse respeito, as teorias de *substrato* e as teorias

<sup>25</sup> Para uma comparação mais detalhada entre esses dois autores sobre a identidade de objetos, ver Krause & Becker Arenhart (2006).

de *pacotes de propriedades*. Ainda que possam haver pequenas variações no âmbito de cada uma delas, as primeiras se assentam na hipótese de que uma entidade como um objeto físico não é apenas constituído pelas suas características (propriedades e relações), como apregoam as segundas, mas que haveria algo para além das propriedades que conferiria a individuação a uma entidade. A natureza desse "algo" permanece em aberto (ver Locke (1995), Livro II, cap.23). Em (2008), Paul Teller afasta essa possibilidade com relação aos objetos quânticos; para ele, no mundo quântico não há substrato ou *haecceities* de qualquer tipo.

Porém, como veremos a seguir, a mecânica quântica nos fornece argumentos para que possamos questionar essa noção intuitiva dada por alguma teoria de pacotes de propriedades (mais abaixo). Portanto, parece surpreendente que ainda hoje, quando vivemos em um tempo em que lógicas divergentes da lógica clássica são comuns, algumas das quais questionando princípios e conceitos atávicos, como o terceiro excluído ou a lei da contradição, ainda haja um taboo relativamente ao conceito de identidade. Para muitos, como já antecipado acima, esse conceito é inquestionável, necessário para toda e qualquer forma de expressão racional (Bueno 2014). Que isso não é assim, tem sido apontado em alguns artigos (ver, especialmente, Krause & Arenhart 2018).<sup>26</sup>

As entidades quânticas como átomos, elétrons e prótons não podem ser inspecionados visualmente, mesmo que para isso utilizemos microscópios muito potentes — tudo o que teremos serão *representações* dessas coisas, assim como temos representações (simulações) de fractais em uma tela de computador, apesar de que fractais são coisas não computáveis. Não devemos nos iludir, portanto, com as pretensas imagens de átomos que encontramos na internet e em revistas de divulgação. Essas entidades nos são dadas pelas leis das teorias (de pacotes de propriedades) que utilizamos. Por exemplo, um elétron é algo que tem massa  $m = 9,1 \times 10^{-28}g$ , carga elétrica  $e = 4,8 \times 10^{-10}e.s.u.$ <sup>27</sup> e spin =  $1/2$  (em unidades de  $\hbar$ ). Aquilo que satisfaz essas condições é um elétron; são características que podemos chamar (em sentido aristotélico) de *essenciais*. Se algo tiver a mesma carga elétrica e o mesmo valor de spin, mas massa igual a  $1,9 \times 10^{-25}g$ , não é mais um elétron, mas um *muon*. Não há elétrons mais gordos ou mais magros; se mudam-se essas características, mudam-se as entidades. Ademais, todos os elétrons têm essas mesmas propriedades, este aqui na ponta do meu dedo e aquele lá em Antares. Sua distinção é dada pela localização, mas mesmo isso apresenta problemas, como apontado em Krause (2018).<sup>28</sup>

Em resumo, o que se passa é o seguinte. Se queremos tratar de um sistema com dois objetos quânticos de mesmo tipo (tendo as mesmas propriedades "essenciais"), como os dois elétrons referidos, devemos fazer isso por meio de uma função que é simétrica ou anti-simétrica relativamente à permutação dos nomes dessas partículas (os rótulos que somos obrigados a utilizar por causa das limitações de nossa linguagem como "partícula 1" e "partícula 2"), como no vetor (2). Se a tal função (dita função de onda) for simétrica, uma troca de "partícula 1" por "partícula 2" e vive-versa não a altera, mas ela troca de

<sup>26</sup> Outros pensadores do início do século XX colocaram questionamentos relativamente a esse conceito, mas não trataremos disso aqui, deixando essa constatação para um trabalho futuro.

<sup>27</sup> ESU indica o sistema de unidades eletrostáticas.

<sup>28</sup> Se formos descrever o elétron que está no meu dedo pela função de onda (vetor de estado)  $\psi_D$  e o que está em Antares por  $\psi_A$ , quando consideramos o sistema conjunto, a função de onda será algo como a função (2), simplificadamente (exceto por constantes)  $\psi_{AB} = \psi_A + \psi_B$ , uma superposição das duas funções. E, nesse caso, não termos mais como discernir entre os elétrons.

sinal se a função for anti-simétrica (como no caso dos elétrons). Isso na verdade não importa, porque a função de onda (chamemo-la de  $\psi$ ) não tem qualquer papel ontológico. Apenas o seu quadrado  $|\psi|^2$  tem significado físico, representando uma probabilidade. Ou seja, não falamos *desse* elétron (por meio de uma função de onda 1) e *daquele* elétron (por meio de uma função de onda 2). O que fazemos para nos referir a ambos os elétrons é usar uma função anti-simétrica (no caso de férmions) que não distingue *esse* sistema *daquele* sistema. Não há identidade no sentido clássico.

**9.1 Observação sobre a mecânica de Bohm.** Como acima nos referimos à mecânica bohmiana, é conveniente falar algo sobre ela. É discutível se a versão da mecânica quântica devida a David Bohm constitui uma formulação alternativa ou se é uma teoria diferente. O fato é que a metafísica associada a essa mecânica é semelhante à *clássica*. Os sistemas físicos mais elementares são uma associação entre uma *partícula* e uma *onda piloto*, que "guia" a partícula. A partícula tem características "clássicas", tendo um comportamento semelhante àquelas descritas pela física clássica: elas têm identidade, trajetórias bem definidas, dentre outras coisas. No experimento das duas fendas, assume-se que a partícula passa por uma das duas fendas e a onda passa por ambas, e é essa onda que produz o fenômeno de interferência. A mecânica bohmiana, ao que tudo indica, é empiricamente equivalente à mecânica quântica "usual", mas isso também é discutível. As partículas, como na física clássica, têm posições bem definidas a cada instante de tempo, e mesmo suas trajetórias podem ser seguidas por meio de *medições fracas*, que não interferem na trajetória das partículas.<sup>29</sup>

O fato é que, como temos visto, não se pode observar a partícula passando por uma fenda; se isso for feito, o padrão de interferência desaparece. Saliente-se que o resultado observado no experimento não depende da teoria considerada. Assim, a descrição acima de que a partícula passa por uma das fendas e a onda por ambas é apenas uma suposição, mais uma, mas a intuitivamente a mais consoante com a teoria de Bohm.

## 10. E a lógica?

Em 1936, G. Birkhoff e J. von Neumann estenderam algumas idéias avançadas por von Neumann em um livro de 1932. Nesse trabalho, os dois advogaram que considerando os estados dos sistemas quânticos como superposições de autovetores de operadores hermitianos sobre um adequado espaço de Hilbert, os quais representam os observáveis, resulta uma álgebra de subespaços desse espaço que fere a distributividade da conjunção (interseção de subespaços) relativamente à união (mais precisamente, ao espaço gerado pela união de subespaços), fazendo com que esses "conectivos" destoem daqueles da lógica clássica; para detalhes, ver o cap. 8 de Jammer (1974), Aerts et al. (2000). A partir da década de 1960, esse campo começou a ser estudado sistematicamente, caracterizando o que ficou denominado de Lógica Quântica (para históricos, ver o capítulo 8 de Jammer (1974), Beltrametti & Cassinelli op.cit., de Ronde et al. 2018).

Muitas lógicas quânticas foram desenvolvidas a partir daí (Dalla Chiara et al. 2004, Svozil 2005, Pavičić & Megill 2009). No entanto, prosseguiram como um estudo paralelo, notadamente nos dias de hoje, quando a área tornou-se quase que totalmente o estudo

<sup>29</sup> Ocorre algo parecido como quando você arremessa uma pedra e vê a sua trajetória no ar; os fótons que incidem sobre a pedra causam tão pouca modificação em sua trajetória que ela pode ser observada. Mas isso já extrapola os nossos objetivos; o leitor interessado pode ver uma interessante defesa da teoria de Bohm, explicando as medições fracas inclusive, em Bricmont 2016, pp.135-6.

da computação quântica, sem que houvesse uma fundação sistemática da teoria quântica tendo uma dessas lógicas como lógica subjacente. Com efeito, se olharmos um livro qualquer de mecânica quântica ou mesmo de suas extensões à mecânica quântica relativista, não encontramos qualquer lógica quântica sendo previamente descrita. A lógica, bem como a matemática, continua sendo a clássica. Isso certamente funciona bem do ponto de vista do físico mas, como estamos vendo, apresenta uma bela quantidade de problemas dos pontos de vista conceitual e filosófico.

Uma tentativa radicalmente diferente foi proposta por Domenech et al. (2008) e (2010). Nesses trabalhos (ver também Krause & Arenhart 2018b), elaborou-se uma formulação da teoria dos espaços de Fock, que fundamentam a teoria quântica das partículas sem interação, em termos de *quase-conjuntos*, uma teoria matemática que permite a consideração de entidades destituídas de individualidade, como parecem ser os objetos quânticos na visão de muitos. Nesse caso, trata-se sem dúvida de uma fundamentação dessa disciplina utilizando-se uma lógica distinta da clássica.

O tema da *lógica da mecânica quântica*, em distinção à *lógica quântica*, permanece atual e relevante. Se formos levar em conta alguma interpretação particular, como aquela que vê os objetos quânticos como entidades destituídas de identidade em sentido clássico, uma nova base matemática seria recomendável. Esse assunto é discutido particularmente em Krause (2018b).

## 11. Paraconsistência

Para não sermos completamente negativos com relação à possibilidade de se utilizar uma lógica paraconsistente em mecânica quântica, salientaremos dois pontos em adição ao que já se disse acima na nota de rodapé 12. Em da Costa e Krause (2006), usamos uma lógica paraconsistente, denominada de *lógica paraclássica* para acomodar o conceito de *complementaridade*, ao qual demos uma interpretação. Não discutiremos esse caso aqui. O que queremos sugerir é que talvez se possa utilizar essa mesma lógica para dar conta de casos como o do gato. Suponha que adotemos a interpretação dos *muitos mundos* (Wallace 2012). De acordo com essa interpretação, *depois* da medição o mundo de parte em dois (ou mais), um no qual o gato está vivo e outro no qual ele está morto. Esses dois mundos *existem*, são *reais*, mas não se comunicam. Ou seja, temos "gato vivo" e "gato morto", mas não a sua conjunção; a lógica paraclássica não é *adjuntiva*, ou seja, de duas fórmulas  $\alpha$  e  $\beta$  nem sempre se pode formar a sua conjunção  $\alpha \wedge \beta$ . Assim, poderíamos ter "gato vivo" e "gato morto", porém não em um mesmo mundo, como já dito na nota referida acima. Adiantaria em algo? Não creio. O importante é que, mesmo nessa situação, *antes* da medição a situação continua exatamente como descrita acima, sem contradição.

## Referências Bibliográficas

- [1] Aerts, D., D' Hondt, E. and Gabora, L. (2000), Why the logical disjunction in quantum logic is not classical. *Foundations of Physics*, 30 (9): 1453-1480.
- [2] Albert, D. Z. (1994), *Quantum Mechanics and Experience*. Harvard Un. Press.
- [3] Arenhart, J. R. B. & Krause, D. (2016), Contradiction, Quantum Mechanics, and the Square of Opposition. *Logique et Analyse* 59 (235): 301-15.
- [4] Arenhart, J. R. B., Bueno, O. and Krause, D. (2019) Making sense of non-individuals in quantum mechanics. In O. Lombardi, Fortin, S., López, C. and Holik, F. (eds.), *Quantum Worlds. Different Perspectives about the Ontology of Quantum Mechanics*. Cambridge: Cambridge Un. Press: 185-204.
- [5] Baggott, J. (2013), *The Quantum Story: A History in 40 Moments*. Oxford: Oxford University Press.
- [6] Béziau, J. -Y. (2006), Paraconsistent logic! *Sorites* 17: 17-25.
- [7] Birkhoff, G. and von Neumann, J. (1936), The logic of quantum mechanics. *Annals of Mathematics* 37: 823-43.
- [8] Bricmont, J. (2016), *Making Sense of Quantum Mechanics*. Springer.
- [9] Bueno, O. (2014), Why identity is fundamental. *American Philosophical Quarterly*, 51, p.325-32.
- [10] Chibeni, S. S. (1997), *Aspectos da Descrição Física da Realidade*. Campinas, Centro de Lógica, Epistemologia e História da Ciência (Coleção CLE 21).
- [11] da Costa, N. C. A. and Krause, D. (2006), The logic of complementarity. In J. van Benthem, G. Heinzmann, M. Rebuschi and H. Visser (eds.), *The Age of Alternative Logics: Assessing Philosophy of Logic and Mathematics Today*. Springer: 103-120.
- [12] da Costa, N. and de Ronde, C., 2013, The Paraconsistent Logic of Quantum Superpositions. *Foundations of Physics* 43, 845-858.
- [13] Dalla Chiara, M. L., Giuntini, R. and Greechie, R. (2004), *Reasoning in Quantum Theory: Sharp and Unsharp Quantum Logics*. Dordrecht: Kluwer Ac. Pu. (Trends in Logic, 22).
- [14] Dirac, P. A. M. (1967), *The Principles of Quantum Mechanics*. Cambridge: Cambridge Un. Press.
- [15] Domenech G., Holik, F. & Krause, D. (2008), Q-spaces and the foundations of quantum mechanics. *Foundations of Physics* 38 (11): 969-994.
- [16] Domenech, G., Holik, F., Kniznik, L. & Krause, D. (2010), No labeling quantum mechanics of indiscernible particles. *International Journal of Theoretical Physics* 49 (12): 3085-3091.
- [17] Fine, A. (1986), *The Shaky Game: Einstein, Realism, and the Quantum Theory*. Chicago: Chicago Un. Press.
- [18] Folse Jr., H. J. (1986), Complementarity and scientific realism. In Weingartner, P. and Dorn, G. (Eds.), *Foundations of Physics: A Selection of Papers Contributed to the Physics Section of the 7th International Congress of Logic, Methodology and Philosophy of Science*. Vienna: Hölder-Pichler-Temsky, pp.93-101.
- [19] French, S. & Krause, D. (2006), *Identity in Physics: A Historical, Philosophical, and Formal Analysis*. Oxford: Oxford Un. Press.
- [20] Gilmore, R. (1998), *Alice no País do Quantum: A Física Quântica ao Alcance de Todos*. Rio de Janeiro: Zahar.
- [21] Griffiths, R. (2011), *Mecânica Quântica*. 2a. ed., São Paulo: Pearson Prentice Hall.
- [22] Hacking, I. (1983), *Representing and Intervening*. Cambridge: Cambridge Un. Press.
- [23] Howard, D. A. (2017), Einstein's Philosophy of Science, *The Stanford Encyclopedia of Philosophy* (Fall 2017 Edition).
- [24] Kanitscheider, B. (1992), Schrödinger's cat and the interpretation of quantum mechanics. In Götschid, J. et al. (eds.), *Erwin's Schrödinger's World View*. Dordrecht: Kluwer Ac. Pu., 41-51.
- [25] Krause, D. (2017), *Tópicos em Ontologia Analítica*. São Paulo: Ed. UNESP.

- 
- [26] Krause, D. (2019), Does Newtonian space provide identity to quantum systems? *Foundations of Physics* 24(2): 197-215.
- [27] Krause, D. (2018b), Quantum mechanics, ontology, and non-reflexive logics. Pré-publicação, <https://arxiv.org/abs/1802.03805>
- [28] Krause, D. and Arenhart, J. R. B. (2016), A logical account of quantum superpositions. In Aerts, D., de Ronde, C., Freytes, H. and Giuntini, R. (eds.), *Probing the Meaning of Quantum Mechanics: Superpositions, Dynamics, Semantics and Identity*. Singapore: World Scientific, pp. 44-59.
- [29] Krause, D. and Arenhart, J. R. B. (2019), Is identity really so fundamental? *Foundations of Science*. 24(1): 51-71.
- [30] Krause, D. & Arenhart, J. R. B. (2018b), Presenting Non-reflexive Quantum Mechanics: Formalism and Metaphysics. *Cadernos de História e Filosofia da Ciência* 4 (2) Jan-Jun. 2016, pp. 59-91.
- [31] Krause, D., Arenhart, J. R. B. and da Costa, N. C. A. (2012), Ontology and the mathematization of the scientific enterprise. *Phainomeon* (Revista de Fenomenologia) 25: 109-130.
- [32] Krause, D. & Becker Arenhart, J. R. (2006), Hume, Schrödinger e a individuação de objetos físicos. *Revista Eletrônica Informação e Cognition* 5 (2): 59- 71.
- [33] Krause, D. and Gracher, K. G. C. (a aparecer), A comparison among three negations: classical, quantum, and paraconsistent: a case study involving possible paraconsistent objects.
- [34] Kumar, M. (2009), *Quantum: Einstein, Bohr and the Great Debate About the Nature of Reality*. London: Icon Books.
- [35] Locke, J. (1995), *An Essay Concerning Human Understanding*. London: Routledge.
- [36] Pavičić, M. and Megill, N. D. (2009), Is quantum logic a logic? In Engesser, K., Gabbay, D. M. and Lehmann, D. *Handbook of Quantum Logics and Quantum Structures*. Amsterdam: Elsevier, 23-47.
- [37] Pessoa Jr. O. (2003), *Conceitos de Física Quântica, Vol. I*. São Paulo: Livraria da Física.
- [38] Pessoa Jr. O. (2006), *Conceitos de Física Quântica, Vol. II*. São Paulo: Livraria da Física
- [39] Priest, G. and Routley, R. (1989), Systems of paraconsistent logic. In Priest, G., Routley, R. And Norman, J. (Eds.), *Paraconsistent Logic: Essays on the Inconsistent*. Munich: Philosophia, pp. 151-186.
- [40] de Ronde, C., Domenech, G. and Freytes, H. (2018), Quantum logic in historical and philosophical perspective. *Internet Encyclopedia of Philosophy* (<https://www.iep.utm.edu/qu-logic/>). Acesso em 20/11/2018.
- [41] Saunders, S. et al. (2010), *Many Worlds? Everett, Quantum Theory, and Reality*. Oxford: Oxford Un. Press.
- [42] Schrödinger, E. (1957), What is an elementary particle? In Schrödinger, E., *Science Theory and Man*. London: George Allen and Unwin Ltd., pp. 193-223.
- [43] Schrödinger, E. (1980), The present situation in quantum mechanics. *Proceedings of the American Philosophical Society* 124: 323-38. (original alemão de 1935).
- [44] Styer, D. F. et al. (2002), Nine formulations of quantum mechanics. *Am. J. Phys.* 70 (3): 288-97.
- [45] Schlosshauer, M. (2004), Decoherence, the measurement problem, and interpretations of quantum mechanics. *Review of Modern Physics* 76 Oct.: 1267-1305.
- [46] Svozil, K. (2005), Quantum logic. A brief outline. ArXiv: quant-ph/9902042v2 31 Jul 2005.
- [47] Teller, P. (1998), Quantum mechanics and haecceities. In Castellani, E. (Ed.), *Interpreting Bodies: Classical and Quantum Objects in Modern Physics*. Princeton: Princeton Un. Press, pp. 114-41.
- [48] Van Fraassen, B. (1974), The labyrinth of quantum logics. In Cohen, R. S. and Wartofsky, M. W. (Eds.), *Logical and Epistemological Studies in Contemporary Physics*. Dordrecht and Boston: D. Reidel (Boston Studies in the Philosophy of Science, 13).
- [49] Wallace, D. (2012), *The Emergent Multiverse: Quantum Theory according to the Everett Interpretation*. Oxford: Oxford University Press.
- [50] Zeilinger, A. (2005), *A Face Oculta da Natureza: O Novo Mundo da Física Quântica*. Rio de Janeiro: Globo.

# THE COMMON PRESENT IN A BLOCK UNIVERSE

Yuri Balashov<sup>1,2</sup>

## Abstract

Our present experiences are strikingly different from past and future ones. Every philosophy of time must explain this difference. It has long been argued that A-theorists can do it better than B-theorists because their explanation is most natural and straightforward: present experiences appear to be special because they are special. I do not wish to dispute one aspect of this advantage. But I contend that the general perception of this debate is seriously incomplete as it tends to conflate two rather different aspects of the phenomenon behind it, the individual and the common dimensions of the present. When they are carefully distinguished and the emerging costs of the A-theories are balanced against their benefits, the advantage disappears.

## 1. Introduction

Our present experiences are strikingly different from the past and future ones. There is a sense in which the former are available to us while the latter are not. Your experience reading this paper is a case in point; some philosophers say that this experience is *available* to you in a way your breakfast earlier this morning is not. Or suppose I am sitting in a concert, listening to the last movement of Tchaikovsky's Sixth Symphony and having a characteristic - intensely sad and hopeless - experience associated with it, which can be usefully contrasted with the serene and uplifting experience from the previous, third movement. Again, the sad experience is available to me in a way the uplifting one is not.<sup>3</sup>

Every philosophy of time must explain these data. It has long been argued that A-theorists can do it better than B-theorists<sup>4</sup> because their explanation is most natural and

---

DOI: 10.21452/LnA\_serie\_n\_v01\_book\_seminario-logica-no-aviao-2013-2018\_yuri-balashov\_p.123-134

<sup>1</sup> Department of Philosophy, University of Georgia. Email: [yuri@uga.edu](mailto:yuri@uga.edu).

<sup>2</sup> My thanks to Jordan Lavenver, Tom Sattig, Jeff Goodmann and the audience at the Pacific APA meetings in Seattle, Washington (April 2017) for very stimulating discussions of the issues raised in the paper.

<sup>3</sup> To say that only the present experiences are "available" to us is just one - and by no means uncontroversial - way of making sense of the vague feeling that they are special. See Skow 2011: §5-6. The controversial part comes from insisting that 'available' means *available simpliciter*, not *available at t*. Alternatively, one could argue that presentness somehow shows up in the *representational content* of experience, or in its *phenomenal character* (or in both). These two ways of conceptualizing the above-noted intuitive feeling seem to me rather implausible and inferior to the first. But taking a stand on this recently debated issue (see Skow 2011, 2015; Prosser 2016) is orthogonal to my agenda in the paper. So my use of the term 'available' in this context is rather non-committal and merely intended to give a handy label to the peculiar sense in which our present experiences are distinguished over past and future ones according to those who think they are so distinguished. Other terms can and have been used for that same purpose.

<sup>4</sup> Here and below I use the terms 'B-theory of time', 'eternalism', and 'block universe' interchangeably to denote the view that all moments of time and their contents enjoy the same metaphysical status. And I take the denial of this view to be the central feature of all 'A-theories of time.' While all A-theories - presentism, the "moving spotlight" theory, and the "growing block" theory - agree in attributing a privileged metaphysical status to the present they disagree on the nature of this privilege (see the main text immediately below). Although the issues considered in the paper affect all A-theories of time, they



straightforward: present experiences appear to be special because they are special. For the presentist, they are special in being real or existent, as opposed to non-real or non-existent. For the moving spotlight theorist, they are special because the spotlight of the present is shining on them. For the growing block theorist, they are special because they lie on the edge of the growing block.

These data constitute a challenge for the B-theorists who recognize no metaphysical difference among present, past, and future. Think of temporal experiences as items attached to our spacetime trajectories in the eternalist block universe spread out in four dimensions. All your experiences - eating breakfast earlier this morning, reading this paper, and many more - are equally yours, as the block universe contains a single you but many experiences, and each of them is eligible to represent your perspective on the universe. And all my temporal experiences, including the sad and the uplifting ones, are equally mine. You believe, however, that you are reading this paper rather than enjoying the breakfast. And I believe that I am having the sad experience, rather than the uplifting one. This belief in the exclusive nature of our present experiences may or may not be veridical, but it needs to be explained. B-theorists have attempted to discharge this task by combining various linguistic and metaphysical resources ranging from the semantical reduction of tense and the indexical analysis of ‘present’ to radically rethinking the ontology of *you* and *I*.<sup>5</sup> While there is disagreement among the B-theoretic approaches as to how much of the above is necessary virtually everyone - the friends as well as the foes of eternalism - agree that there is some cost to be paid.<sup>6</sup> The general perception is that B-theorists have work to do, perhaps hard work, to accommodate the apparent distinction between present and non-present temporal experiences, and that explanation does not come about so naturally as in the A-theories. As a result, the A-theories have enjoyed significant dialectical advantage on this particular score.<sup>7</sup>

I do not wish to dispute one aspect of this advantage. But I contend that the general perception of this local debate is seriously incomplete and may be distorted, as it tends to conflate two rather different aspects of the phenomenon behind it: the *individual* and the *common* dimensions of the present. When they are carefully distinguished and the emerging costs of the A-theories are balanced against their benefits the advantage disappears. Or so I argue below.

## 2. Individual and Common Present

The above examples mention our individual present experiences (IP) - your reading this paper and my listening to the last movement of Tchaikovsky’s Sixth Symphony. These experiences seem exceptional to you and me, but nothing in the block universe favors them over our corresponding past or future experiences. Although this poses a *prima facie* problem for B-theories let us set it aside here and assume that somehow or

---

can be illustrated most clearly in the context of presentism, which is the main focus of my discussion; but I will make a few comments on other A-theories as well.

<sup>5</sup> For a classical B-theoretic defense of the semantical reduction of tense and the indexical analysis of ‘present’, see Mellor 1998. For arguments that these resources are not enough, and that explaining the exclusive nature of the present experiences requires explicit rejection of the common-sense view that persons persist by enduring, in favor of stage theory, or perdurance, see Hoy 1978; Balashov 2005, 2015, 2017; Torre 2010; and Skow 2011, 2015.

<sup>6</sup> Important exceptions include Mellor (1998: 42-45) who thinks that no cost is to be paid, that the problem of the “presence of experience” (his name for it) is easy for the B-theorist.

<sup>7</sup> Which may, of course, be offset by the disadvantages afflicting the A-theories on other scores, from physics to truth-making. I put them aside here.

other - perhaps by developing a convincing account of the indexical nature of IP<sup>8</sup> or by explicitly adopting the stage view of persistence - B-theorists can handle this problem and resist the A-theoretic notion that explaining IP requires a unique and objective present. It is still true that every person has many temporally restricted perspectives on the block universe, each confined to a particular time. But let us suppose that this does not stand in the way of explaining why our ‘current’ individual perspectives appear so exceptional to each of us.

I submit, however, that this is only part of what needs to be explained. The other part has to do with the fact that we all seem to share a common present. As aptly put by Butterfield, “we think of ourselves as sharing a common, albeit ever-changing, now” (1984: 161), where this is explicated as follows: “two people share a now if they agree in their judgments about what is now the case, i.e. in their present-tensed judgments” (1984: 174). This characterization will be made more precise below. But to give an example, suppose you and I are at the concert together and, shaken by the experience of listening to the final movement of Tchaikovsky’s Sixth Symphony, I whisper to your ear, “It’s so sad!” You whisper back, “Yes, it’s hopeless.” Clearly, our agreement is about short stretches of our individual experiences separated by a few seconds. On a much larger scale, we also agree - in a different sense of ‘agree’ - that 2019 is present, that Trump’s presidency is present, and so forth. Such agreement can manifest itself in many direct and indirect ways, but initially it seems to pose a separate problem for the eternalist. Even if she can explain away our beliefs in the exclusive nature of our IPs there seems to be a further problem of explaining why they are “coupled” and “coordinated” in ways that allow us to share a common present. For example, my 21:00:00 symphony-listening stage should be thus coupled with your 21:00:10 stage<sup>9</sup> but not with your 20:45:00 stage (which is exposed to the previous, serene and uplifting, movement). Why are they so coupled? Call this the problem of the common present (CP).

As I understand it, my notion of CP (which will be further detailed in the next section) is very close to what Butterfield (1984) and Callender (2008) call a “common now.” They both think a common now is something on which we can “agree” or “disagree,” and I used this language in my initial description of CP above. But this use is rather special and needs a brief comment or two. First, we can disagree about a lot of things for a thousand reasons, simply because we may have different beliefs about something. We need to set all of that aside. The only agreement or disagreement of interest to us has to do with whether something is going on, or is present. Second, such agreement or disagreement is normally implicit. It manifests itself in various remarks, if we make them; in our responses to certain stimuli and prompts, if we do respond to them; and so forth. This suggests that it may be helpful to think of it as having a broadly dispositional nature. Assuming all that, how can we possibly “disagree” about the present - about what is currently going on, or about “what is now the case” (as Butterfield puts it)?

---

<sup>8</sup> From now on ‘IP’ will be interpreted more broadly to include not only individual experiences (such as seeing a green traffic light), but also individual actions (such as crossing a road) and other episodes and events, both physical and mental, involving individuals.

<sup>9</sup> To allow for a short time lag involved in our quick communication and in perception processing. More on such time lags below. For the purpose of this discussion I use ‘stage’ as a neutral term not committing one to a particular view of persistence. Thus ‘my 21:00:00 symphony-listening stage’ is roughly synonymous with ‘the 21:00:00 symphony-listening temporal stage of my life’. This is how the term was used by many philosophers (e.g. in debates about personal identity) before the recent development of the stage theory of persistence, where the term acquired a more demanding meaning.

I can think of a few cases. Suppose you are watching a basketball game and listening to a commentary, and it all goes very fast. The commentator says, “Wow, isn’t that great!”, referring to Michael Jordan’s pass to Dennis Rodman who then scores a fantastic slam dunk, which causes you to say or think, a split second later, “That is great!” Clearly, you and the commentator do not share a common present, because you react to different external events, which rapidly succeed each other. You do not disagree on the quality of Jordan’s pass, but you fail to agree about its presentness.

Here is another case, on the opposite side of the spectrum. Suppose you and I communicate by email, and we both think that everything Obama does is wonderful. But I am really slow in responding to emails. You write, “Our president is wonderful, as always!” A few weeks later I get to read your email and respond (referring to Trump), “No, he is not.”

These examples are not straightforward and may eventually turn on the peculiar referential behavior of indexicals and demonstratives. Why is it not straightforward to come up with examples of disagreement on the common present, on whether something is going on? I think it is because we have been conditioned by evolution to agree on it, not because it is impossible to disagree. The time lags involved in perceptual processing, the duration of the relevant external events, the way our communication goes, our reaction times, and in general the speed with which we tend to respond to various stimuli and prompts have all been shaped and selected by evolution to maximize our success. Our coherent perception of the common present<sup>10</sup> - our implicit agreement on “what is the case” - is part of this success.

To illustrate this point, consider another case. Imagine you and I are in a jungle and there is suddenly a tiger in front of us. Suppose your relevant dispositional state, including your readiness to do certain things and avoid others, can be associated with your belief that there is a tiger in front of us, and that we should do something to save our lives. And suppose my dispositional state can be associated with my belief that there is no tiger in front of us, although there could be. In this situation, we may “disagree,” in the above-noted extended sense of the term, on whether something is going on. And it may have dramatic consequences. It also explains why we rarely in fact disagree about CP in such cases.<sup>11</sup>

Characterizing these cases in terms of disagreement about a common present requires walking a fine line: we do not want an interesting concept of CP to rule out the possibility of disagreement about it; but we do want to maintain that we agree about it most of the time. Alternatively, one could say that in such cases we do not “disagree” about a common present; rather we simply *lack* a common present of a certain kind. And instead of saying that we “agree” on a certain CP, we could say that we *share* or *have* it.<sup>12</sup> The problem of CP could then be put as follows: why is it that some of our IPs (e.g. my breakfast and your breakfast) can be parts of a certain CP (namely, the kind of CP that is normally associated with having meal together), but others (e.g. my breakfast and your dinner) cannot?

---

<sup>10</sup> More precisely, of the *many* common presents; see the next section.

<sup>11</sup> As Callender puts it, “large time lags plus fast tigers wouldn’t help reproduction” (2008: 348).

<sup>12</sup> This alternative terminology eliminates the distraction of working with the special sense of “agree” and “disagree” noted above. For stylistic reasons, and to preserve some continuity with the earlier literature, I shall occasionally use the terms *agree* and *agreement* below when talking about *sharing* a CP, or simply *having* a CP. As indicated above, agreement or disagreement in question is most often implicit, may involve broadly dispositional states, and concerns only the temporal characteristics of the relevant external events and of our IPs.

When stated in these terms the problem may initially strike one as being completely trivial. And in one sense, this is entirely correct (see §4-6 below). But it is important not to lose track of where we are in our discussion at this point. We began with the B-theoretic picture of IPs scattered across the Block Universe. We then assumed that the B-theorist could explain why each of us has a particular privileged IP. Again, some think this is easy - perhaps “as easy as indexicals,” to put it in a slogan. Others disagree. But we presumed that this could be done, one way or another, and that our IPs emerged from this picture as being *locally privileged*. So far so good. But I think we cannot stop here. Suppose my breakfast is locally privileged for me, and your dinner is locally privileged for you. Why do these IPs *not* belong to a common present of the relevant sort? I submit this is an interesting further question to raise, at this particular juncture.

We tend to pass over CP because it is not such an immediate - obvious and striking - datum as IP. Your current individual pleasure or pain imposes itself on you in a way that is hard to ignore. Our realization that we share or fail to share a common present is, in contrast, at a remove from immediate experience. Directing attention to it requires a shift of focus and some abstraction. And it may remain indirect even after that, lurking in some dispositional form. More importantly, it requires interaction and communication, and it involves time lags (of which more below). But despite being a bit elusive, CP is a significant fact to be accounted for. I contend that the ability to account for it is no less important in the overall cost-benefit analysis than the ability to explain IP.

Now, it might seem that the A-theories have an obvious advantage here too; for couldn't they explain IP and CP in the same uniform way by relating it to the objective ontological global present? And CP might initially look like more headache for the B-theorist; she had to work (hard) to make sense of the apparent exclusive nature of our IPs, and now she seems to be saddled with an additional problem of pairing them up in the right way.

Surprisingly (or perhaps not; see below), the situation is reversed here. The eternalist can offer a very elegant, compelling, and almost automatic explanation of CP. Indeed it is so compelling that the best strategy for the A-theorists of all stripes is simply to borrow and adapt it to their metaphysics of time. I consider this option in §5. I will argue, however, that in doing so, the A-theorists incur significant costs that undermine the unity of their position and reduce the benefits they may enjoy from IP.

### 3. The Nature of the Common Present

But the first order of business is to get clearer on the nature of CP. As already noted, our realization that we share a common present is most often implicit and indirect. We rarely ask each other questions about what is currently going on. Most of the time we simply act, communicate and interact with one another and with our environment in many different ways. CP is best thought of as a by-product of this ongoing interaction. So unlike IP, CP is not an individual experience or action, nor even a sum of such items. CP is partly *constituted* by IPs, but normally involves a lot more. If you and I are exposed to some external event then both the event and the relevant stretches of our IPs - those involved in our interactions with the event and with each other - become part of our common present, as we would be quick to confirm if asked. In most cases, our interaction with each other and with our environment constitutes a complex network of entangled processes including, at the very least, the external events themselves, our reactions to them, our communication with each other, as well as other related events, processes, and actions, which may be long and short, direct and mediated, active and passive, uni- and

bidirectional, with various degrees of overlap and no well-defined boundaries. It is not possible to describe the infinite variety of such transactions. But taken together, they create a sense of a common present in us.

More precisely, each sufficiently delineated subnetwork of interrelated external events, the relevant stretches of our IPs, and our communication about them circumscribes a specific common present. The notion of CP so understood is *causal, temporally-extended, context-relative, vague, and local*. Each of these features requires a brief comment.

First, CP is *causal*. This is obvious from the above description.

Second, in addition to being extended in space, CP is *extended in time*. This is a consequence of its causal nature. Suppose you and I hear a roaring thunder and realize that it occurs in our common present. The CP that we thus share includes, at a minimum, certain characteristic sound wave disturbances in our vicinity, our auditory responses to them, which themselves take time, various reactions that may result from it, as well as our subsequent communication about our individual experiences.<sup>13</sup> Besides spanning a certain region of space (say, a few thousand feet) these transactions span a finite interval of time (say, two or three seconds). Together they occupy a local spacetime region. My individual experience of the thunder, your experience of it, as well as our exchange about our experiences are all part of this CP. A worthwhile contrast here is with your experience of the lightning ten seconds earlier. The lightning is not part of the relevant common present.

On the other hand, suppose we exchange views about Trump's first presidency. Even a short fragment of our exchange may take quite some time and span a longer common present. Trump's ongoing presidency is part of this CP whose extent may range from several minutes to several days or even months (if we communicate by email or on blogs). The contrast here may be with Obama's presidency, which is no part of this CP.

This shows, third, that the temporal extent of CP is *context-relative*. What we take to be our common present may be short or long depending on the context. More examples: sharing a CP in the context of listening to a symphony movement requires exposure to roughly the same relatively short sequence of musical sounds. My 21:00:00 stage and your 21:00:10 stage are in this situation. Failure to be so exposed constitutes failure to share a CP. My 21:00:00 stage and your 20:45:00 pm stage are in that situation. On the other hand, living during a war sets a rather different margin for sharing a CP. Our stages separated by 15 min will normally fit in it, but our stages separated by a month may not.

The above three features of CP have a consequence that each of us can at any time belong to many different CPs of different extents and different degrees of overlap with each other.

Fourth, CP is typically *vague* in that its spatial and temporal boundaries are not precisely defined, even in a given context. Many factors contribute to this vagueness.

---

<sup>13</sup> The time lags normally involved in perception and communication do not prevent us from forming a solid notion of a common present in ordinary cases of observing and touching nearby objects, or hearing sounds produced by nearby sources, or communicating with nearby persons. In most such cases the experiences we communicate to each other (and which we take to be occurring in our common present) are caused by objects and processes that change sufficiently slowly for our reports of them to be veridical and to fit properly in the margins created by the time lags of communication, observation, and so forth. The significance of such time lags in the genesis of our idea of the present has been discussed by Butterfield (1984), Stein (1991), and Callender (2008). And where the changes are very fast it is often unclear whether our observations and reports refer to the same common present. In such cases we cannot be sure that we really share a common present when communicating with each other; recall the basketball commentary example from §2.

But as with other vague phenomena, there are clear cases of falling within a CP (in a given context) and clear cases of failure to do so. The above examples provide some such cases. Our experiences of the thunder, as well as our quick communication about them, are clearly part of the CP associated with the thunder. Your experience of the lightning is clearly not part of it. My experience of the very first sound of the thunder may be a borderline case. And all of the above are clearly part of our CP associated with Trump's presidency.

Fifth, CP is *local*. Its spatial extent is limited to the events that can have some, perhaps indirect, causal influence on us, and to those we can have influence on, during the finite stretch of time associated with a given CP. What is happening in the Magellanic Clouds right now (if the question even makes physical sense) is not part of our common present (i.e. no part of any of our CPs). This immediately explains why we do not care about it. In general, it is natural for us to care about what is within our "causal reach" and ignore what lies outside it. And when it comes to CP it is natural for us to restrict its scope to the events that belong to a causal nexus of processes directly involving or at least partially engaging our IPs. The details may be complicated and vary considerably from case to case (the contextual nature of CP comes into play here). But limiting the scope of CP to local occurrences explains why we are inclined to include in our common present a certain roaring thunder, but not a certain process, however dramatic, happening right now in Io's atmosphere.<sup>14</sup>

There is, no doubt, more to say about the nature of CP. But what has been said is enough to revisit and put into sharper focus the questions posed at the end of the previous section: Why do we share many common presents? Why does a given CP couple our IPs in the right way? What is the *coupling mechanism* whereby our individual, locally privileged IPs become entangled in a common present?

#### 4. The Common Present and Self-Selection

First, put yourself in the shoes of an eternalist who believes in a single block universe filled, among other things, with our worldlines extending from birth to death. Various short fragments of such lines incorporate our IPs. Some of them interact with each other and with the local events, and form the notion of CP, as described above. Why do we generally "agree" on what it includes? That is to say, why do we *share* a certain context-relative CP? The eternalist explanation is embarrassingly simple. Indeed, it is difficult to pose the above question without answering it. We agree on, or share, a particular CP because the agreeing (sharing) can only be done by those of our locally privileged IPs that belong to this CP.

As is clear from the description in the previous section, the common present is a causal phenomenon. Our "agreement" on a given contextually-relevant CP - our coming to *share* it - is just another causal process implicated in it. To agree on a given CP, we need to interact with each other and with the local events. And causal interaction of this sort naturally produces agreement thanks to a self-selection effect involved in the relevant causal transactions. Why do my 21:00:00 stage and your 21:00:10 stage agree that the music is so sad? Because the particular CP associated with it includes the short stretches of our IPs borne by our individual stages, which *also* do the agreeing, within the

---

<sup>14</sup> Such a process should not be confused with our perception of it, which may be happening right now and be part of our CP. The local nature of the common present, or "common now," has been emphasized by Stein (1991), Butterfield (1984), Gibson and Pooley (2006), and Callender (2008).

relevant time margins involved in auditory processing and subsequent communication.<sup>15</sup> In contrast, your 20:45:00 stage and my 21:00:00 stage cannot share a musical CP of the relevant sort (associated with the last - sad and hopeless - movement of the symphony), and hence cannot be constitutive of the corresponding CP, simply because they do not belong to a common causal network of processes typically associated with fleeting musical experiences.<sup>16</sup>

The eternalist explanation of CP is thus very efficient: CP emerges as a by-product of ongoing causal interaction of the subjects of experience amongst themselves and with their environment. It originates in the self-selection of locally privileged IPs spread over the block universe. Self-selection has plenty of material to work on in the block universe! Not so in the presentist universe. Or at least not obviously so.

The presentist universe considered at a certain time includes, say, my 21:00:00 concert experience and your 21:00:00 concert experience. It does not, in addition, include your 20:45:00 concert experience. Nor does it include your 21:00:10 experience. How does a particular CP get built from the IPs available to the presentist? If the presentist were to be guided by her native ontology she should start by thinking of CP as “IP writ large.” To wit, she should insist on the *ontological* nature of the IP-coupling mechanism: our individual IPs become part of a single CP by sharing an objective global present, and they do so in virtue of their sheer existence. If this explanatory strategy worked it would provide an elegant unified account of both IP and CP. But it doesn’t work. The presentist may have a good story to tell about IP, by grounding it in the global ontological present. But the latter is at best an idle wheel in explaining CP, which has a *causal* and not an ontological origin.

To see it even more clearly, try this thought experiment. Posit the presentist’s global ontological present but eliminate causal interaction among the subjects of experience, say, by confining them to isolated places; think of the Monte Cristo story.<sup>17</sup> Each of them would still have a sense of a distinguished, locally privileged individual present, and the presentist could relate it to the global ontological present.<sup>18</sup> But the subjects would not have a sense of a *common* present. And to the very limited extent they did - say, by gleaning incidental bits and pieces of information from prison guards, hearing occasional external sounds, catching vague glimpses of light and shadow through the window, and so forth - their limited sense of the common present (i.e. of the present they share with their immediate environment) would be *causal* from top to bottom. The presentist’s global ontological present would be of no help in instilling it in them.

Now eliminate the objective ontological present but restore causal interaction and populate the universe with the rest of the eternalist material (i.e. the past and the future).

<sup>15</sup> See notes 9 and 13.

<sup>16</sup> This does not prevent them from sharing another, more generic kind of CP associated, say, with the whole symphony.

<sup>17</sup> Edmond Dantès, the protagonist of Alexandre Dumas’s famous novel *The Count of Monte Cristo*, was imprisoned for many years in the Château d’If, lost track of time, and only regained it when he met a fellow prisoner, Abbé Faria.

<sup>18</sup> How? Presumably, by saying that the present experiences located in the global ontological present are all the experiences there are. The moving spotlight theorist will say that the present experiences are those in the spotlight. The growing block theorist will say that the present experiences are those on the edge of the growing block. This sense of the privileged present instant is crucial to the A-theories’ respective accounts of the distinguished *individual* present, for it is expected to explain why subjects have a privileged sense of what they are experiencing now in the first place; it must explain the *presence* of their individual experiences. Different A-theories approach this explanatory task in different ways. I cannot go into details here.

The eternalist would then have to confront the problem of the individual present and come up with some admittedly revisionary account of it. That is to say, she would need to reconcile the existence of her multiple IPs from different times with her belief in the privileged status of one of them, which she calls *the* present. But it is precisely the abundance of IPs from different times that also gives her a very natural way of building a common present from the multitude of individual presents by means of causal interaction and self-selection. Indeed, the eternalist does not need to “build” it. All she needs to do is watch it being self-built: the only temporal stages of you and me that can agree that they live in a common present are those that do live in a common present, since sharing a CP requires causal overlap. Adding the objective global present could neither add nor subtract anything from this account. This shows that the global present plays no role in it.<sup>19</sup>

One might object that this reasoning confuses the *existence* of the common present with our *perception* or *concept* of it, and that the A-theorist still has a better explanation of the former. On her view, the common present exists out there, objectively, and we simply become aware of it by interacting with each other and with other objects belonging to the common present, and by making use of short-term memory. The eternalist cannot tell this kind of story about CP.

But is it a good story? What we become aware of when we form a concept of CP is not the A-theoretic present. Our concept is of the *causal, temporally extended, and spatially local* common present. Consider distant objects such as the Magellanic Clouds. They are every bit as real as we are, and what goes on in them right now (again, if the question makes physical sense) is part of the A-theoretic present. But is it part of our common present? No. Our concept of CP is *local*, while the A-theoretic ontological present is global. So if one wants to explain the origin of our CP then the eternalist’s causal network of local processes, contextually bound into a particular CP, must be the centerpiece of such explanation. It does all the heavy lifting.<sup>20</sup>

To sum up: the common present has a causal nature and is grounded in the self-selection of the individual stages of experience. The eternalist can simply take this at face value and offer a very natural explanation of CP, which draws on the plenitude of times and of the individual temporal experiences available in a block universe.

I think this calls for a reassessment of the dialectical situation between eternalism and its rivals *vis-à-vis* the present. Eternalism may be revisionary when it comes to explaining the distinguished status of the locally privileged individual present experiences. But A-theorists cannot extend their native accounts of individual present experiences to the common present.

## 5. Discussion

But why should they? If the above is correct - if the common present is causal and not ontological - then what prevents the A-theorists from simply appropriating the causal

<sup>19</sup> Even if it does play a central role in explaining the subjects’ privileged access to their individual present experience. This point applies to all A-theories, not just to presentism. See note 18.

<sup>20</sup> Can’t the A-theorist reasonably ask whether something (or someone) in the Magellanic Clouds exists right now? If so, isn’t she getting at the notion of common present? No. If she is a presentist, she is getting at the notion of *existence*, not of common present. If she is a moving-spotlight or a growing-block theorist, she may be getting at the notion of “lustrous existence” - a metaphor that can be developed into theories by non-presentist A-theorists. For a recent critical overview of several ways of connecting the moving spotlight of the objective present to the subjective experience of the present, see Skow 2015, which also provides useful references to earlier discussions.



account of it and adapting it to their metaphysics of time? After all, eternalism has no monopoly over causation, and A-theorists of all stripes must be able to make sense of causal relations anyway. This is certainly true, and A-theorists have risen up to that challenge. But it is a well-known challenge. The cost of meeting it is especially high for presentism. Being a cross-temporal relation, causation has its natural place in the eternalist universe. The presentist, who denies the existence of more than one time, has work to do to accommodate such relations, perhaps by grounding them in presently instantiated tensed properties, or in *abstracta*.<sup>21</sup> While this is no place to debate the merits and demerits of such proposals there is no doubt that they involve some cost. Causation and other cross-temporal relations are initially foreign to the presentist ontology and necessitate decidedly revisionist moves, specifically designed to recreate ersatz analogs of non-present times and their contents.<sup>22</sup> In terms of the dialectic of the debate between the rival theories of time, the most important point here is that *after* such adjustments are made, it becomes clear that the main work of explaining the common present is done by the elements appropriated from the eternalist framework, where they occur naturally, not by the distinctive elements of the A-theoretic ontology.

So after the reassessment, the overall dialectical situation seems to be as follows. The B-theory may be initially unfit to deal with the locally privileged individual present experiences. To make sense of IP, the eternalist needs to do some work (or so some argue). But once this work is done, CP emerges automatically. The A-theorists, on the other hand, have an easy time with IP. But the A-theoretic notion of IP is of no use in explaining CP. To make sense of the latter, the A-theorists need to borrow the essential resources from the rival theory and pay some costs<sup>23</sup> for adapting these resources to their respective ontologies.

In addition, by treating CP and IP as equally *non-ontological*, the eternalist position emerges as more unified in one respect. In contrast, the A-theories are, at the end of the day, committed to the view that IP and CP are entirely different in nature, the former being ontological and the latter causal.

## 6. Conclusion

Debates about the status of the present in the philosophy of time involve examples such as our current pains or pleasures and someone's present perception of a green light

<sup>21</sup> For details of such proposals see, for example, Bigelow 1996 and Caplan and Sanson 2011.

<sup>22</sup> The moving-spotlight and the growing-block theories may be in a better position here, as they do not deny the existence of more than one time; so there seems to be room for bona fide cross-temporal relations. I need to confine myself to the barest outline here, but things are not so simple. Consider a cross-temporal causal relation between two items  $e_1$  and  $e_2$  located at times  $t_1$  and  $t_2$  respectively. Unlike the eternalist, the moving-spotlight theorist who wants to develop a general account of causal relations cannot discharge the bulk of her explanatory task simply by pointing to these items. She needs to say that what makes them "animated" or causally effective is their presence in the moving spotlight. (This becomes particularly important in explaining the "lustrousness" of the present temporal experiences; see Skow 2011 and 2015: Ch. 11 for a useful discussion.) So she needs to say that  $e_1$  was causally effective (as a cause) when the spotlight was shining at it; and that  $e_2$  was (or is) causally effective (as an effect) when the spotlight was (or is) shining at it. But the spotlight cannot be shining at both times. More carefully, there is a sense in which it cannot, and this sense creates a predicament similar to the presentist's predicament. Perhaps this sense can be explained away, or shown to constitute no insurmountable problems for an account of cross-temporal causal relations. But some work needs to be done. *Mutatis mutandis* for the growing-block theory.

<sup>23</sup> Ranging from moderate (in the case of the moving-spotlight and growing-block theories) to significant (in the case of presentism).

or of a sad musical passage; but also examples such as collective concert experiences, presidential elections and terms, and wars. For instance, speaking of the Second World War, an A-theorist may claim that it is objectively past, that the moving spotlight has moved away from it, or that it is already deep inside the growing block. All of that is fair game. But phenomena of the second sort may require a separate treatment. They are not anyone's individual experiences; however, they often involve such experiences. The theory needs to explain what makes a certain war present, not just for a single person, but for many people.

My discussion has taken up this task. It has assumed that the B-theorist can adequately explain why each of us has a locally privileged individual present - think of my breakfast and your dinner - and moved past this problem to a different question: what in the Block Universe disqualifies these IPs from sharing a common present of the relevant kind - the kind normally associated with having meal together? It may appear, initially, that the A-theorist already has the beginning of an answer to this question based on her account of the "local privilege." All she needs to do is simply extrapolate it to the common present. But this turns out to be a non-starter, for the common present is not assembled from IPs ontologically, rather it is self-selected from them causally. And it is the B-theorist who has the beginnings of a good account along these lines, as he embraces the whole plenitude of IPs on which self-selection can act. His ontological inventory includes my breakfast, as well as yours; but also your dinner; and a lot more. Nothing needs to be done to "assemble" a common present from them; we can simply watch it being self-assembled. This is a fascinating show, and the A-theorists need to buy into it. But they have to pay the familiar cost of recreating non-present times and their contents.

To recap, the original dialectical situation between the A- and B-theories is as follows: they agree that they both need to explain some phenomenon, IP, which is naturally at home in the A-theories, but initially foreign to the B-theory; so the B-theory faces a battle. The dialectical situation after the reassessment is different: there are *two* phenomena to explain, IP and CP. They are different in nature, but both need to be accounted for. IP is naturally at home in the A-theories, but foreign to the B-theory. But CP is naturally at home in the B-theory, and initially foreign to the A-theories, which need to pay some price - perhaps a familiar price - for dealing with CP.

I conclude that the general perception that the A-theories have an obvious edge over the B-theory when it comes to explaining the present is incorrect. This task has two equally important parts. The A-theories may fare better on the first, but lose to the B-theory on the second.<sup>24</sup>

---

<sup>24</sup> This leaves some interesting questions unanswered. We moved too quickly past the initial problem of IP. This was done on behalf of those B-theorists who feel that they have nothing to fear from it. But the problem deserves further discussion. Knowing more about the B-theoretic account of IP would help to better understand how the theory deals with CP. More intriguingly, couldn't the potential of the B-theory to explain CP cast light on how it should deal with IP in the first place? That would be an interesting kind of "bootstrap explanation." I said earlier that it may be natural for some A-theorists to think of CP as "IP writ large"? But couldn't we instead think of IP as "CP writ small"? I cannot address these questions here.

## Bibliography

- [1] Balashov, Yuri (2005), "Times of Our Lives: Negotiating the Presence of Experience," *American Philosophical Quarterly* 42: 295-309.
- [2] Balashov, Yuri (2015), "Experiencing the Present," *Epistemology and Philosophy of Science* 44: 61-73.
- [3] Balashov, Yuri (2017), "Time, Fission, Fusion: An Argument against the Block Universe with Endurance," *Manuscripto* 40 (1): 109-143.
- [4] Bigelow, John (1996), "Presentism and Properties," *Philosophical Perspectives* 10: 35-52.
- [5] Butterfield, Jeremy (1984), "Seeing the Present," *Mind* 93: 161-76.
- [6] Callender, Craig (2008), "The Common Now," *Philosophical Issues* 18: 339-61.
- [7] Caplan, Ben and David Sanson (2011), "Presentism and Truthmaking," *Philosophy Compass* 6 (3): 196-208.
- [8] Gibson, Ian, and Oliver Pooley (2006), "Relativistic Persistence," *Philosophical Perspectives* 20: 157-198.
- [9] Hoy, Ronald (1978), "Becoming and Persons," *Philosophical Studies* 34: 269-280.
- [10] Mellor, D.H. (1998), *Real Time II*. London and New York: Routledge.
- [11] Prosser, Simon (2016), *Experiencing Time*. Oxford: Oxford University Press.
- [12] Skow, Bradford (2011), "Experience and the Passage of Time," *Philosophical Perspectives* 25: 359-387.
- [13] Skow, Bradford (2015), *Objective Becoming*. Oxford: Oxford University Press.
- [14] Stein, Howard (1991), "On Relativity Theory and Openness of the Future," *Philosophy of Science* 58: 147-167.
- [15] Torre, Stephan (2010), "Tense, Timely Action and Self-Ascription," *Philosophy and Phenomenological Research* 80: 112-132.

# CHANGE AND CONTRADICTION: A CRITICISM OF THE HEGELIAN ACCOUNT OF MOTION

Emiliano Boccardi<sup>1</sup>

## Abstract

In his *In Contradiction* (1987), Priest levelled three powerful arguments against the received Russellian view of change and motion. He argued that his preferred paraconsistent theory of change, the Hegelian account, is immune from these objections. Here I argue that these three arguments are sound, but that the Hegelian account falls pray to them too. I conclude, however, that the Hegelian account is in a better position to tackle these challenges.

## 1. Introduction: the received view of motion and change

The idea that things change and that time passes is deeply entrenched in our cognitive and emotional lives. Yet, as soon as we try to explicate these notions in clear conceptual terms, we are immediately faced with great difficulties. Of all the ways in which this worry can be expressed, perhaps the most famous, and the most perspicuous for our purposes, are Zeno's celebrated arguments against motion.

Consider an object in uniform motion from location  $x$  to location  $y$  (the tip of Zeno's arrow, for example). And consider any particular instant during its journey,  $t_0$ , during which the tip is located exactly in the position  $x_0$ , and only at that position. At, or during  $t_0$ , the arrow advances not at all away from  $x_0$ , in its journey to  $y$ . Any such advancement, in fact, would take a finite amount of time, however small. The duration involved in this advancement, therefore, would not be an instant, contrary to the hypothesis. Now, the whole interval of time during which the journey of the object supposedly occurs is made up of such instants. Since no advancement is achieved during any of those constituting instants, it seems to follow that no advancement can be made in any number of them, however great this may be. If no advancement can be made during any time interval, then no advancement is made by the object, contrary to the hypothesis.

At any time  $t_0$  during its journey, according to this premise, the tip of the arrow is located exactly in the position  $x_0$ , and only at that position. As Aristotle put it in his famous reconstruction, at any time during its journey, the arrow "occupies a space equal to its size". What makes this premise nearly compelling is that denying it seems contradictory. How could the tip of the arrow be and at the same time be not located at position  $x_0$ ? How could the arrow occupy a space greater than its size?

According to an often-recounted philosophical gossip, one that is reported in any introductory physics textbook, this dilemma was finally solved by the invention of calculus. As they struggled to come up with a suitable mathematical description of continuity and continuous change, Leibnitz and Newton, independently from one another, developed the fundamental notions that constitute what we now know as calculus. Despite its immense fruitfulness in the development of modern science, and its wide spread use among physicists and mathematicians, it was not until the end of the 19th Century that these

---

<sup>1</sup> Department of Philosophy, Universidade Federal da Bahia. E-mail: [emiliano.boccardi@gmail.com](mailto:emiliano.boccardi@gmail.com)  
DOI: 10.21452/LnA\_serie\_n\_v01\_book\_seminario-logica-no-aviao-2013-2018\_emiliano-boccardi\_p.135-148

notions found a coherent systematization in what is known as the “classical foundation of calculus”. As we shall see, however, such coherentization was obtained at the price of a radical reconceptualization of the relevant notions, one that arguably deprives them of their intuitive appeal in the dissolution of Zeno’s paradoxes. The central idea involved in this reconceptualization is the quantificational understanding of limits: rather than treating derivatives as ratios of infinitesimal quantities, according to this (now familiar) understanding, limits should be understood as quantifiers.

Let us consider how this idea applies to the notion of instantaneous velocity. Instantaneous velocity, according to the classical view, should be understood as the limit of a sequence of ratios of finite quantities:

$$v(t_0) = \frac{dx}{dt}(t_0) = \lim_{\Delta t \rightarrow 0} \frac{x(t_0 + \Delta t) - x(t_0)}{\Delta t}$$

This formula, in turn, should be understood as expressing the fact that the values of the finite ratios  $\frac{x(t_0 + \Delta t) - x(t_0)}{\Delta t}$  become indefinitely close to  $v(t_0)$ , the limit, as we chose smaller and smaller values of  $\Delta t$ : for all  $\epsilon > 0$ , there exists a  $\delta > 0$  such that for all  $\Delta t < \delta$ :  $|\frac{x(t_0 + \Delta t) - x(t_0)}{\Delta t} - v(t_0)| < \epsilon$ .

It should be stressed that the expression “become indefinitely close” ought to be thought of as a dead metaphor, reminiscent of a time when the continuum was conceived as the product of actual motion, and variables as denoting varying quantities: “infinitesimals are an attempt to extend to the values of a variable the variability which belongs to it alone. When once it is firmly realized that all the values of a variable are constants, it becomes easy to see, by taking any two such values, that their difference is always finite, and hence that there are no infinitesimal differences.” ([17], p. 357).

Russell enthusiastically took the classical coherentization of calculus as providing us with a coherent response to Zeno’s challenge: “people used to think that when a thing changes, it must be in a state of change, and when a thing moves, it is in a state of motion. This is now known to be a mistake. When a body moves, all that can be said is that it is in one place at one time and in another at another.” ([16], p. 69).

This admittedly paradoxical idea that things could change without ever being changing, also known as the at-at theory of change, appears to be uniquely capable of making sense of change without postulating true contradictions, and quickly became the received view.<sup>2</sup>

## 2. Priest's three objections to the Russellian view

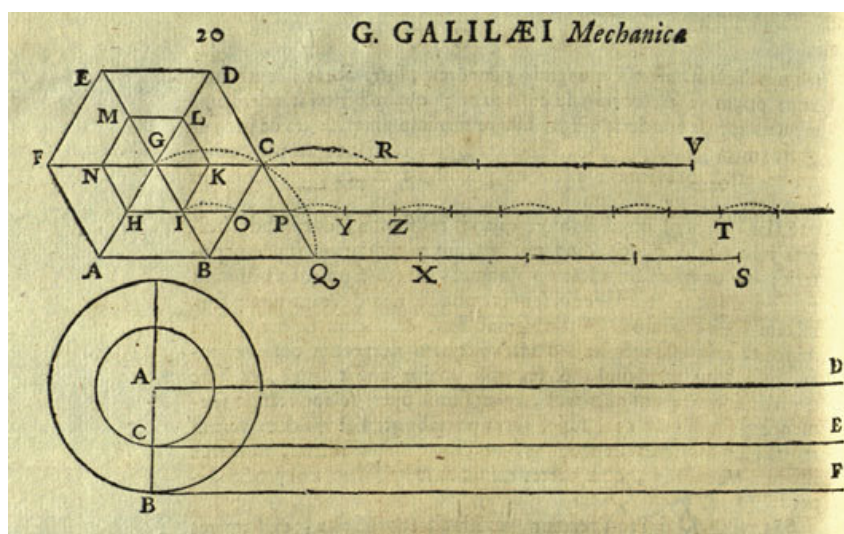
In his book *In Contradiction* ([14]), Priest put forward three lines of criticism to the at-at theory of change, with which I fully agree. In this section I will present these criticisms and defend them from some possible objections.

<sup>2</sup> This idea was not new, and a debate over its virtues can be traced back at least to medieval times. William of Ockham and his followers advocated a view according to which change is nothing over and above a sequence of different properties had at different times. It was known as the doctrine of changing form (*forma fluens*), and it opposed the dual-fact account known as the doctrine of change of form (*fluxa formae*). According to the latter, when a leaf (for example) changes its color from green to yellow, not only does it have a determinate shade of (say) green at any given time, but, over and above, it also has a changing shade of green (the dual fact). According to the former, on the contrary, there is no further fact over and above the leaf having different shades of green at different times.

**2.1 The problem of incremental accretion.** As we have seen, the source of the discomfort one feels when confronted with the arrow paradox is that one finds it hard to intuit how the arrow could manage to get anywhere by a series of going-no-where's. Part of the appeal of the original (incoherent) understanding of infinitesimals was precisely that it appears to alleviate this discomfort, since it allows us to conceive of any stretch of time or space as composed of infinitesimal 'line elements', whose individual contribution to the total advancement of the arrow, while infinitesimal, is not null.

A way to illustrate the intuitive appeal of the original conceptualization is to contrast the case of a polygon revolving on a plane without sliding, with the case of a circle so advancing (see fig. 1).<sup>3</sup>

As the polygon advances, each of its sides touches the plane in its due turn, so that when a total rotation is completed the total stretch of space touched by the polygon, equal in size to its perimeter, can be seen to consist of a finite sum of discrete "jumps", each equal in size to one of its sides. If we try to replicate the same reasoning to the case of a



circle revolving on the plane, we are instantly met with a great difficulty. Here too, the figure touches the ground with one element at a time, each in its due turn. However, unlike what happens in the case of the polygon, the element of the circle that touches the ground at each time is a dimensionless point. Since there is no sliding, moreover, just as it happens in the case of the polygon, the velocity of each of these point-elements relative to the ground is a 0.

How could the circle manage to advance, if its advancement consisted of a series of null contributions, each of which in its turn consists of a dimensionless point of the circle touching the ground without moving with respect to it? Prima facie, construing of the circle as a limit case of a polygon with infinitely many "sides" or "line elements", partly alleviates this discomfort, allowing us to describe the two situations as (at least partly) analogous. But this understanding of continuous motion is precisely what the classical foundation of calculus is supposed have banished:

These days no one worth his salt thinks instants "add up to" periods this way. If there are instants, periods are instants with distance-relations between them. The relations, not their relata, account for periods' extension:

<sup>3</sup> This paradoxical scenario, known as the problem of the Rota Aristotelis, was first discussed in the Greek work *Mechanica*, traditionally attributed to Aristotle. Upon completing a full rotation, the centre of the greater wheel will be displaced by a distance equal to its circumference. However, as the greater circle completes a full rotation, the smaller wheel, which is attached to it, completes a full rotation too. So, by an analogous reasoning, we ought to conclude that the centre of the smaller wheel (which is the same as that of the greater one), will have moved by a distance equal to its smaller circumference. A contradiction. The scenario was famously discussed by Galileo in his *Two New Sciences*.

that is why (in the paradox) without putting distance-relations between the points, we don't get an extension. [Leftow 2014 [12], p. 239].

However, the mere definition of a mathematical function, called "distance", whose values are totally independent from the individual advancements achieved at each instant during the journey, can at best be seen as a restatement of the problem, rather than its solution. As Priest rightly points out:

That one can prove a small mathematical theorem or two is one thing; but it does not ease the discomfort that one finds (or at least, that I find) when one tries to understand what is going on physically, when one tries to understand how the arrow actually achieves its motion. At any point in its motion it advances not at all. Yet in some apparently magical way, in a collection of these it advances. Now a sum of nothings, even infinitely many nothings, is nothing. So how does it do it? [ibid., p. 175]

The problem we are discussing should be distinguished from the mere problem of understanding how a static infinite aggregate of dimensionless elements could 'add up' to a finite quantity. In the dynamic setting that we are considering, this problem is further aggravated by the fact that these elements enter the scene one at the time, so that it is hard to understand how we could deny that these solitary facts should make a "contribution" to the total feat. In the case of a static, or standing variety of infinity, all the elements are already there, waiting to be named and counted, as it were, and so are all the relations that obtain between them. Two points of a line have the distance they have quite independently from the existence of any intervening points. These standing elements need not be *reached* in any sense, nor should their distance count as an *achievement* of any kind. The situation is crucially different in the case of a growing variety of infinity, which is what is at issue here. To be told that the distance between the bow and the target is some finite quantity only proves that *if* the target could be reached, then the total advancement of the arrow would be equal to that quantity. It tells us nothing, however, about whether such feat can be accomplished, since it tells us nothing about the physical formation of this path during the journey. To use Bergson's apt words, according to the Russellian view, "movement is composed of immobilities" ([4], p. 308).

**2.2 The problem of the non-intrinsicity of motion.** Standard physics textbooks tell us that the state of the universe at any time is determined by the state at any previous time plus the laws of physics. The state of the universe at a time  $t$ , we are told, is given by the positions of all the particles at  $t$ , plus their velocities at  $t$ . However, according to the Russellian view, the instantaneous velocities of the particles supervene solely on their positions at various times preceding and succeeding  $t$ , hence they add nothing to the state of the universe at exactly the time  $t$ . Now, the positions of particles at  $t$  have clearly no effect whatsoever on the positions in the immediate future of  $t$ . How could their velocities have any greater explanatory role, if they supervene solely on these positions themselves? As Priest put it:

It follows from the definition that there is no such thing as an intrinsic state of motion. If one had a body in motion and took, as it were, a logical "picture" of it at an instant, the picture obtained would be no different from one of the same body at the same place, but at rest. Of course, an object in motion can have an instantaneous non-zero velocity, but it would be wrong to think that this differentiates it intrinsically from a static body. [ibid., p. 173]

Some authors tried to derive less dramatic consequences from the application of calculus in response to Zeno's argument, denying that the quantificational understanding of limits and derivatives renders instantaneous velocity a relational property (cf. Smith 2003 [18]). I have argued elsewhere that this view, known as "the modified at-at theory of change", is untenable.<sup>4</sup> Here I shall content myself with presenting a symptom which afflicts the Russellian account, and which I think lends indirect support to Priest's observation.

It has been often noted that it is hard to understand how Russellian instantaneous velocities could fulfill the explanatory role that is standardly ascribed to them in physics. This is well illustrated by the following argument by Frank Arntzenius:

Consider a ball moving from right to left through some region in space, and a qualitatively identical ball, perhaps the same one, some time later, moving from left to right through that very same region. If the full state at a time of a ball does not include an instantaneous velocity then the full state of the two balls is exactly the same when they occupy the same region. [...] Why does the one ball subsequently move to the left and the other ball subsequently move to the right? Surely it is conservation of velocity, or something like that, which determines that the one ball will keep moving to the left and the other will keep moving to the right. But if there is no such thing as instantaneous velocity, as there is not in the at-at theory, then why do the balls continue their motions in different directions? [[2]: § 2.]

This objection has been raised several times under different guises in the recent literature.<sup>5</sup> It appears to expose a fatal shortcoming of the received view, one that derives directly from the non-intrinsicity of change. The general problem appears to be that Russellian velocities make the state of the universe at an instant conceptually dependent on its states at other instants, since the velocity objects have at a time is a property of their trajectory around that time, not an instantaneous intrinsic property.<sup>6</sup> This deprives these "states" of all their explanatory power.

**2.3 The problem of the direction of time.** The third and last criticism raised by Priest against the standard view is that it appears to be incapable of accounting for the intrinsic directedness of time and change, the so-called anisotropy of time. Time and change always happen in one direction only, from earlier to later, from past to future. Here then is the difficulty, in Priest's words:

What accounts for the anisotropy of time? This again has been a thorny problem, particularly for those who have denied the reality of the flow of time. They have had to locate the anisotropy of time not in time itself, but in processes in time - a tall order, since apparently all causal laws are time-symmetric. [Ibid., p. 116].

Now, while Priest is surely right that reconciling the apparent irreversibility of physical processes with the time-reversal invariance of fundamental laws is a "thorny problem", I think that mentioning this in that context obscures the full force of his objection. To see why, it is worth discussing briefly a common fallacy. It has often been suggested that the second law of thermodynamics, which implies that the entropy of a closed system never decreases, could be used to explain and reduce the directionality of time itself. According

<sup>4</sup> [6], pp. 164-168.

<sup>5</sup> See for example [1], [11], [2], [19], [5], [9].

<sup>6</sup> See for example [1], p. 10.



to these suggestions, the future direction of time is nothing but the direction in which global entropy increases. The passage of time would be grounded by the fact that the entropy gradient is not zero. But these reductive accounts cannot succeed. The second law of thermodynamics, in fact, can only establish a contingent correlation between the entropy gradient and the direction of time. If the intrinsic direction of time was not fixed independently from the entropy gradient, how could we even meaningfully express the empirical content of the law? What does it even mean to say that entropy never “increases”, if not that it never increases in the direction of the future?

Imagine a world where the second law does not hold. It would be a world where broken glasses would spontaneously recombine to end up on the tables from which they fell. But in order to even represent this possibility, we have to assume that the broken-glass states in that world would be instantiated *before* the on-the-table states. If the direction of time itself were reversed in such a scenario, then that world would be just indistinguishable from ours! In fact, the second law of thermodynamics would hold in that world too, and by necessity, contrary to the hypothesis. As Tim Maudlin quite aptly put it, if the future-directed arrow of time amounted to nothing but the direction in which entropy increases, “[entropy] could go up and down like the stock market, but since the ‘direction of time’ would obligingly flip along with the entropy changes, entropy would still never decrease” (2007, p. 129).

What was said about the second law applies equally well to any other time-asymmetric law postulated for grounding the direction of time, regardless of whether it be a fundamental law or a statistical one (as is the case with the second law). The general lesson to be learned by the failure of these reductive attempts is this. Mere asymmetric relations are never enough, by themselves, to fix a direction. At best, asymmetric relations can determine an order among the elements of their domains. But mere order is “blind” to directionality, so to speak. Three points on a line, for example, have an order. B lies between A and C, etc. However, these betweenness relations are insensitive to the direction in which the series might be transversed (from A to C, or from C to A). Asymmetric relations determine an order, but not a direction. Unfortunately, this direction, or sense, is precisely the feature that any account of time and change should explain, and that the Russellian account struggles to accommodate.

### 3. Priest's Hegelian account of motion

This is where we stand. Zeno's arguments appear to show that time and change are contradictory. The main attempt to remove these contradictions without denying time and change altogether, the now widely accepted Russellian account, proves to be at best a restatement of the problem. We are thus faced with a dilemma. Either accept the phenomenological datum that things change and time passes, at the cost of accepting that at least some contradictions are true; or hold fast to the principle of contradiction, and repudiate time and change as mere appearances. Heraclitus and Hegel (according to some interpretations) opted for the first horn of this dilemma. Parmenides, Zeno and McTaggart, instead, opted for the second one.

To make matters worse, as I have argued in my Boccardi and Perelda 2018 ([8]), the claim that time and change are mere appearances runs the risk of being itself self-contradictory, lest it be carefully qualified. In claiming that our experience represents the world as instantiating dynamic features absent from reality, the denier of change must be careful not to concede too much, as she would if her account required that dynamic experiences themselves (or their contents) instantiate dynamic properties incompatible

with the B-theory. It is not enough to claim that passage is mind dependent in the sense that it requires the interaction of a conscious subject with external reality (as is the case, for example, in [3]). The strict denier of passage wants to further claim that the content of the experience of passage is never true of the actual world, not even of that part of it consisting of the interactions between subjects and their objects: 'the dynamic character of our immediate experience does not require time itself to be dynamic' ([10]: p. 391).]

This view, I argue, is self-undermining. If we concede that change and time are mere appearances, doesn't it follow necessarily from this that at least these appearances really change? If every time that things appear to us to change we make a mistake, in that the things that appear to change don't really change, isn't it still true that we make different such mistakes at different times, and hence that at least appearances themselves really do change?

Everything suggests that time and change are real and contradictory. Very few philosophers, however, took seriously this option, since questioning the principle of contradiction threatens to blow a fatal strike at the heart of rational thinking itself. It is unfortunate that the few philosophers who took up the challenge of denying the principle, most notably Heraclitus and Hegel, are among the most obscure and cryptic thinkers of all times. The philosopher Graham Priest, in more recent times, devoted a good part of his career to the herculean task of raising the vague contention that (some) contradictions might be true to the highest standards of clarity and analytic thinking.

This unprecedented accomplishment was made possible, among other things, by the discovery and development of logical systems that block the Principle of Explosion (ex contradictione quodlibet). These systems, known as paraconsistent logics, opened the way for a non-trivial treatment of true contradictions, or dialetheia. One paraconsistent logic that is particularly apt for this purpose is the so-called logic of paradox (LP), introduced by Priest in his [13]. Using this logic as background, Priest was able to put forward a dialetheist account of motion that promises to overcome the three deficiencies of the standard account that we discussed above.

Let me introduce this account. I will then proceed to illustrate how the account is supposed to overcome the three difficulties mentioned above. Finally, I shall argue that the Hegelian account suffers from the same difficulties as the Russellian account (sec. 4).

According to Priest, what prevented us from developing a suitable treatment of time and motion, vis a vis the problems that we have been discussing, is the assumption that bodies in motion occupy coherently one exact position at the time, at the expense of all others:

Consider a body in motion—say, a point particle. At a certain instant of time,  $t$ , it occupies a certain point of space,  $x'$ , and, since it is there, it is not anywhere else. But now consider a time very, very close to  $t$ ,  $t'$ . Let us suppose that over such small intervals of time as that between  $t$  and  $t'$  it is impossible to localize a body. Thus, the body is equally at the place it occupies at  $t'$ ,  $x'$  (Hence, at this instant the body is both at  $x$  and at  $x'$  and, equally, not at either). [Ibid., p. 176]

The idea can be spelled out as follows. Let the motion of the body be represented by the equation  $x = f(t)$ . According to the Russellian view, the proposition that the body is located at position  $r$  at time  $t$  is true if and only if  $r = f(t)$ . It is false if and only if  $r \neq f(t)$ . According to Priest's account, instead, for each time  $t$  during the motion of the body, there is an interval containing  $t$ ,  $\theta_t$ , called a "spread", such that, if  $t' \in \theta_t$ , the

occupation of the body's location at  $t'$  is "reproduced" at  $t$ . In other words, the proposition that the body is located at position  $r$  at time  $t$  is true if and only if, for some  $t' \in \theta_t$ ,  $r = f(t')$ . It is false if and only if for some  $t' \in \theta_t$ ,  $r \neq f(t')$ .

To see why this view is dialetheist, it is perhaps better to characterize it in terms of state descriptions. Let  $q$  denote any quantity, ranging over a set of values  $V$  (for example, the real numbers). Let  $v(t) = q$  be a function of time, and  $Q(r)$  be the predicate that  $q$  has value  $r$ . Finally, let  $S_t = \{Q(r) : r = v(t)\} \cup \{Q(r) : r \neq v(t)\}$ . The propositions contained in  $S_t$  state that at  $t$  the value of  $q$  is  $v(t)$  and not any other value.  $S_t$  can then be thought of as the state description of the system at time  $t$ . The spread hypothesis amounts to the claim that, for every time  $t$ , there is an interval containing  $t$ ,  $\theta_t$ , such that the complete state description at  $t$  is the set theoretic union of all the consistent state descriptions  $S_{(t')}$  for all the  $t' \in \theta_t : \cup_{(t' \in \theta_t)} S_{(t')}$ . Clearly, if the state of the system is changing in a neighbourhood of  $t$ , such union will be inconsistent.

**3.1 Priest's account of temporal passage.** One of the most interesting features of Priest's treatment of change is that it promises to be applicable to all sorts of change, including the peculiar kind of change - if it is a kind of change at all - which the passage of time consists of.

When change is conceptualized in a Russellian fashion, as the instantiation of different states at different times, it becomes conceptually absurd to claim that time itself changes in any sense. And yet there is something impermanent about time. We make a gesture towards this feature when we refer to the incessant advancement of the present towards the future. It is perhaps not a chance that all the metaphors that we use to express the impermanence of time are metaphors of motion. We say that time passes, that it flows, that it flies. We approach the end of the semester, etc. What does the passage of time have in common with change in general, and with motion in particular? What is the cash value of these metaphors?

In this regard, the intrinsicity of change in Priest's account constitutes a great advantage. We have discussed the account as it applies to motion, but we might as well have applied it to any other magnitude as well, charge, mass, or whatever. According to the Hegelian account, for a quantity, any quantity, to be in a state of change is for it to be in a certain contradictory state.<sup>7</sup> This very general idea of change lends itself immediately to an application to the case of time itself. Here is how this is supposed to work.

Let  $q$  denote the state of time, whatever this may be, and  $V$  the set of the reals. The proposition  $Q(t)$ , for example, could mean that time  $t$  is present. Let  $v(t)$ , finally, be the identity function. Applying the spread principle as we did above, one can see that, for each time  $t$ , there exists a interval  $\theta_t$  such that, for all  $t' \in \theta_t$ ,  $Q(t')$  is true at  $t$ . The claim that time is passing, or that the time that is present is constantly "changing", can then be seen to rest on the fact that the spread  $\theta_t$  is non degenerate, or equivalently that the state description of time is always inconsistent.

Let us now turn to Priest's application of the Hegelian account to the three challenges presented above. In what follows I illustrate the proposed application of the Hegelian account to tackle the three challenges discussed above. I argue that the account falls pray to all of them.<sup>8</sup>

<sup>7</sup> Priest of course is not denying that if a quantity assumes different values at different times then it has changed. The claim is rather that this is not what change could consist of.

<sup>8</sup> Some of the arguments against the Hegelian account presented here can be found in my Boccardi and Bustos 2017 ([7])

#### 4. The Hegelian account and Priest's three objections

**4.1 How the Hegelian account is supposed to tackle the problem of incremental accretion, and why it fails.** Priest argues that the Hegelian account of motion ought to be preferred to its rival, the standard Russellian view, since the spread hypothesis solves at one stroke all the difficulties mentioned above. Let us see how this is supposed to work in each case, starting from the problem of incremental accretion. The problem, remember, is that of explaining how a series of non-advancements might produce a finite displacement, a positive advancement. Here is how the spread hypothesis is supposed to overcome this difficulty:

The Hegelian account of motion may be taken to locate a fault in the [arrow] argument, but at a point different from that upon which Russell lights. For, according to Zeno's argument, at a particular point in time the object occupies only a single point in space, whence it follows that it advances not on its journey during that instant, i.e. that the measure of the set of points occupied at that instant is zero. Given the spread hypothesis, however, it is not true that the moving body occupies only a single point. At an instant,  $t$ , it occupies all the points in  $\Sigma_t$ , which is, in general, not a singleton. Indeed, provided the function of motion,  $f$ , is continuous,  $\Sigma_t$  is an interval, and therefore has non-zero measure. Thus, advance is made during a single instant, and hence during the aggregate of instants. [Ibid., p. 180]

This response, I argue, squeezes out of the Hegelian account more than it can deliver. The unwarranted assumption, crucial for giving the appearance of a solution to Zeno's paradox, is that the mere fact that the instantaneous location of the object be of non-zero measure would suffice to make of it an "advancement". A way to expose the role of this subtle but crucial assumption is to note that Zeno's argument can be easily generalized to apply to the Hegelian account as well. Here is a Hegelian friendly version of Zeno's arrow argument.

At each moment during its journey, the arrow occupies (inconsistently) a region of space the same size as its spread  $\Sigma_t$ . During each instant  $t$ , the spread  $\Sigma_t$  of locations occupied by the arrow advances not at all. Then how does it manage to advance over a finite interval of time, given that this is constituted by nothing but a sequence of such spreads?

The inconsistency of the Hegelian locational schema resides in the fact that the spread  $\Sigma_t$  which the arrow occupies at time  $t$  is larger than the size of the arrow itself. But why should we think that this larger space occupied by the arrow constitutes an advancement of the arrow? If at each moment during the journey the arrow occupies the spread  $\Sigma_t$ , then the motion of the arrow can only consist of its subsequently occupying different spreads, each in its due turn. How this feat is achieved, the account does not say.

Another way to expose this difficulty is by considering Priest's intriguing hypothesis that quantum mechanical indeterminacies might constitute the physical ground of the spreads:

Perhaps the measure of  $\Sigma_t$ ,  $\sigma(\Sigma_t)$ , just is the uncertainty in the location of the object at  $t$ . Perhaps quantum mechanical indeterminacies are fundamentally the result of inconsistencies in motion, and in particular in the spread postulated by the spread hypothesis. This suggestion at least allows us to give physical significance to the spread. (Ibid.)

The difficulty under discussion here can be exposed by constructing a quantum version of Zeno's arrow, and then testing the Hegelian interpretation against it. Consider then a particle whose locational state at time  $t$  is given by the wave function  $\Psi(x, t)$ . The Zenonian argument then goes as follows. Over the finite interval of time  $[t_0, t_1]$ , the state of the particle changes from  $\Psi(x, t_0)$  to  $\Psi(x, t_1)$ . But at each time during its journey the state of the particle changes not at all. How is it, then, that in a sum of changing-not-at-all's the particle manages to instantiate a different state?<sup>9</sup>

On the Hegelian interpretation, the instantaneous states represented by  $\Psi(x, t)$  should be construed as representing inconsistent superpositions of exact locations. But notice that the inconsistency of the account is never used by Priest to solve the problem of incremental accretion. As Priest noted in objecting to my criticism, the only feature of it that is relevant is the non-zero length of the location occupied by the particle at the instant.<sup>10</sup> However, it should be noted that whether the particle occupies a single dimensionless point at an instant, as the standard view has it, or an inconsistent spread of locations, it does so consistently: on the Russellian account the particle occupies a point of space at the expense of all others, while on the Hegelian account it occupies a spread of locations at the expense of all others. The challenge is to explain how the location occupied might change without ever being changing. It seems to me that with regard to this, the Hegelian account faces exactly the same difficulty as the Russellian one.

Of course, the discomfort would be alleviated if the view was that the arrow proceeds by "jumping" discretely from one spread of locations to the adjacent one. In this case, its advancement would be analogous to that of the revolving polygon, where each side touches the ground all at once, contributing to the whole journey with an advancement equal to its finite length. But this is not Priest's view. The edges of the spread, as well as all the other points inconsistently occupied by the tip of the arrow at an instant, are all durationless points. During each instant of time, all these locations advance not at all. Yet, if the arrow is to reach its target, it has to occupy all of these points, one by one, in succession. How does it do it?

One can also express this worry by noting that the whole arrow, unlike its tip, *does* occupy a finite amount of space at each instant during its journey. Yet, nobody is tempted to consider the extension of the arrow as an "advancement" towards its target. The spreads occupied by the tip of the arrow at each instant of time, I argue, are no different from the whole arrow, when it comes to accounting for the advancement of the arrow. Each position inconsistently occupied by the tip of the arrow has no element of movement, just like the positions consistently occupied by it in the Russellian account don't.

**4.2 How the Hegelian account is supposed to tackle the problem of non-intrinsicity of motion, and why it fails.** As we have seen, one of the problems with the standard account is that it makes intrinsic velocities - and states of change in general - relational or neighborhood properties. The Hegelian account promises to solve this problem rather well:

This is obviously no problem for the Hegelian account. For it, there is an intrinsic state of motion: a certain inconsistent state. The difference between a body genuinely in motion and one changing place but at rest

<sup>9</sup> In my Boccardi and Bustos 2017 ([7]) we argue that the quantum interpretation of the spread hypothesis is problematic, for a technical reason that I shall not repeat here.

<sup>10</sup> See Priest's response at minute 41:30 of <https://www.youtube.com/watch?v=iR58cnYyviw&t=2014s>.

each instant is exactly that between a Hegelian state description and the corresponding Russellian one. [ibid., p. 180]

Priest is surely right that his account renders motion an intrinsic feature of objects. And this is *prima facie* an advantage. The problem, remember, was that Russellian intrinsic velocities are properties of trajectories, rather than intrinsic temporary properties of objects. This rules them out as potential explanations for these trajectories themselves. Now, the positions that the arrow may (inconsistently) occupy at any given time according to the Hegelian account are totally unrelated to the locations it may occupy at any other time. So, the Hegelian account at least does not have this problem.

However, it could be argued that the inconsistent spread of locations at a time is not the right kind of intrinsic state. Notice, in fact, that in a Russellian world the positions (consistently) occupied by all objects at a time do constitute a temporary intrinsic state of the universe at that time. Yet, nobody would be tempted to consider that state as the ontological ground of instantaneous motion. Why? Presumably, because the position of an object at a time is totally, not merely conceptually, independent from its position at other times. Two identical locational states of the universe may yet diverge radically as to their dynamic properties.

Now, how is the inconsistent spread of positions postulated by the Hegelian account different in this respect? The spread occupied by the arrow at a time is totally, not merely conceptually, independent from the spread it occupies at other times. So why should we think that it has anything to do with motion? Why should the instantiation of inconsistent states of affairs as to the location of the ball at a time “force” it, or “induce it” or “dispose it” to be elsewhere at other times?

It will be objected that this criticism asks too much from a theory of motion. After all, the complaint against the Russellian view was that it renders the velocities of objects at a time  $t$  (and hence the instantaneous states of the universe at that time) conceptually dependent on the positions of the objects at all times around  $t$ , thus depriving them of their explanatory power. Isn't it preposterous, then, to complain that on Priest's account intrinsic states of motion do not entail the correlated trajectories? The spread hypothesis, by itself, does not explain the trajectories precisely because it complies with the requirement that it be conceptually independent from them! At best, one could require that there be a causal, or law-like connection between intrinsic states of change and trajectories. This is exactly what Priest does (Ibid., p. 179).

I think that this line of defense is fair. However, I still think that my objection points at a difficulty with the Hegelian account, for two reasons. First, the ambition of the theory is to provide an account of change in general, not of motion only. We have seen, for example, how this generality allows us to apply the Hegelian account to the passage of time itself. If Priest were right, true contradictions would not just contingently correlate with trajectories: they would constitute the ontological ground of change in general. They would be the essence of change, as it were. This is one of the chief allures of the Hegelian account. Given the generality of the account, however, I think it is not unreasonable to demand an elucidation of this systematic connection. Of course, one could postulate a different causal connection for each kind of change. However, the general pattern of connections would still cry for an explanation, lest it be a bizarre fluke.

The second worry that I have with the “just so” response that we are considering, is that it is in tension with Priest's use of the spread hypothesis in tackling the problem of incremental accretion. The problem, as James put it, was that the “conceived positions, however numerous multiplied, contain no element of movement, so Zeno, using nothing

but them in his discussion, has no alternative but to say that our intellect repudiates motion as a non-reality." [James 1987: 735.]. If the spread hypothesis is to tackle this problem, the contradictory states must not only provide us with a cause for the trajectories. They must further be the basic instantaneous elements of motion itself. Each spread occupied during the journey of the arrow must make a contribution to the whole journey. Therefore, each spread must be, and not merely cause, a kind of displacement. If things are so, the total lack of conceptual dependence between contradictory state descriptions and trajectories in the Hegelian account is just as problematic as the total conceptual dependence between them in the Russellian account.

**4.3 How the Hegelian account is supposed to tackle the problem of the direction of time, and why it fails.** It is undoubtedly an interesting advantage of Priest's treatment, as we have seen, that it applies to the passage of time just as well as it applies to any other kind of change. Priest also claims that the Hegelian view can account for the intrinsic directionality of passage, something that standard accounts are at great difficulty to explain. On this latter point, I dissent from Priest. The solution proposed rests on the observation that: " $\theta_t$  will not necessarily be distributed symmetrically about  $t$ . In fact, there are reasons to suppose that  $t$  is the leading edge of  $\theta_t$ , so that the interval is skewed all to the past of  $t$ , at least normally." [Ibid., p. 214]

This asymmetry in the distribution of the spread, according to Priest, would solve the problem of grounding the directionality of time:

Again, the solution to the problem on the present approach is obvious. Since the identity function is monotonically increasing,  $v(t)$  is always the upper bound of the spread of  $q$  at  $t$ . Thus, the direction of the flow of time is perpetually from past to future, which seems just about right. [ibid.]

The problem with this solution, just as for the Russellian account, is that it tries to ground directionality on mere asymmetry. We have already seen how this feat is impossible. Priest surreptitiously avails himself of a primitive notion of directedness when he claims that the identity function is monotonically "increasing". The identity function, by itself, is neither increasing nor decreasing, just like the great pyramid of Giza is neither ascending nor descending.

The problem can be aptly illustrated by considering the order induced on the real numbers by the relation smaller-than. Like any relation that is asymmetric, connected and transitive, the relation smaller-than induces an order on the reals. In addition, the reals present a structural asymmetry. This can be seen by noting that the positive reals have a square root, while the negatives don't. As Reichenbach ([15], p. 26) noted, this asymmetry can be exploited to differentiate structurally the relation smaller-than from its converse, the relation greater-than. This can be done, for example, by saying that any number which is the square of some number is greater than any number which is not the square of some number. This feature of the reals provides the series with a kind of "directionality". However, this anemic kind of directionality is not the kind of directionality that Priest is seeking to ground. Mere order, even embellished by the perfection of structural asymmetry, does not determine any directionality. To see this, it suffices to observe that, despite the structural asymmetry of the relation smaller-than, it simply makes no sense to ask whether the reals are increasing or decreasing, or whether they "run" in one direction as opposed to the other.

## 5. Conclusion

Zeno's arguments appear to show that time and change are contradictory. The main attempt to remove these contradictions without denying time and change altogether, the now widely accepted Russellian account, proves to be at best a restatement of the problem. We are thus faced with a dilemma. Either accept the phenomenological datum that things change and time passes, at the cost of accepting that at least some contradictions are true; or hold fast to the principle of contradiction, and repudiate time and change as mere appearances. Priest has shown that the first option is not as abhorrent as we may think, and that it ought to be seriously explored. He also claimed that the Hegelian account ought to be preferred to the Russellian one on the ground that it is immune from the problems of incremental accretion, of the non-intrinsicity of motion, and of the directionality of time. I have argued that this is not the case. The Hegelian account does not help overcoming these difficulties.

However, I think that Priest's account does present a number of advantages over the Russellian one. The main one is its generality. As we have seen (sec. 3.1), the account can be applied to the case of temporal passage, a problem that is apparently intractable within a Russellian world view. Secondly, even if the spread hypothesis does not solve the three problems, it seems to be in a better position to solve them. For example, even if the hypothesis does not solve the problem related to the non-intrinsicity of change, at least it points in the right direction, by making change an intrinsic state of affairs. For the same reason, I think that Priest's view is also in a better position than its rival when it comes to solving the problem of the directionality of time. As we have seen, the mere asymmetry in the spread could not, by itself, account for the directionality of time. However, whatever does so ground directionality, arguably, must at least supervene on some intrinsic state of affairs.

Finally, Priest's account helps to explain why the issue of contradiction proved so recalcitrant in connection with time and change. Contradictions, qua contradictions, I have argued, don't appear to contain any essentially dynamic element. However, the suspicion that the idea of change might be contradictory has continued to resurface over and over in the history of philosophy, and all attempts to dissolve these paradoxes proved to be unsatisfactory. Thus, it should not come as a surprise if it turned out that change essentially involves true contradictions.

**Acknowledgment.** I am very grateful to Graham Priest for his incisive comments. I am also indebted to Moises Macias Bustos, with whom I co-authored a paper which contains some of the ideas discussed here. Finally, I wish to thank the editors Rodrigo Freire, Alexandre Costa-Leite and Edgar Almeida.



## Bibliography

- [1] D. Z. Albert. *Time and Chance*. Cambridge (MA), London: Harvard University Press., London, 2000.
- [2] F. Arntzenius. Are there really instantaneous velocities? *The Monist*, 83:187–208, 2000.
- [3] L. R. Baker. Temporal reality. In M. OâRourke J. K. Campbell and H. S. Silverstein, editors, *Time and Identity*, pages 27–47. The MIT Press, Cambridge, MA, 2010.
- [4] H. Bergson. *Creative Evolution*. Holt, Reinhart and Winston, New York, 1911.
- [5] J. Bigelow and R. Pargetter. *Science and Necessity*. Cambridge University Press, Cambridge, 1990.
- [6] E. Boccardi. Time as motion. *Metaphysica: International Journal for Ontology and Metaphysics*, 19(1):1–31, 2018.
- [7] E. Boccardi and M. Bustos. Contradictions in motions: Why they’re not needed and why they wouldn’t help. *Humana.Mente, Journal of Philosophical Studies*, 32:195–227, 2017.
- [8] E. Boccardi and F. Perelda. The delusive illusion of passage. *Analysis*, 78: 3:387–396, 2018.
- [9] J. Carroll. Instantaneous motion. *Philosophical Studies*, 110:49–67, 2002.
- [10] B. Dainton. Time, passage, and immediate experience. In C. Callender, editor, *The Oxford Handbook of Philosophy of Time*, pages 382–419. Oxford University Press, Oxford, 2011.
- [11] M. Lange. How can instantaneous velocity fulfill its causal role? *The Philosophical Review*, 114 (4):433–468, 2005.
- [12] B. Leftow. Instants, events, and god. In N. Oaklander, editor, *Debates in the Metaphysics of Time*. Bloomsbury, London, 2014.
- [13] G. Priest. The logic of paradox. *Journal of Philosophical Logic*, 8(1):219–241, 1979.
- [14] G. Priest. *In Contradiction: A Study of the Tranconsistent*. Clarendon Press, Oxford, 1987.
- [15] H. Reichenbach. *The Direction of Time*. Dover Publications, Mineola, NY, 1956.
- [16] B. Russell. Mathematics and the metaphysicians. In B. Russell, editor, *Mysticism and Logic*. Barnes and Noble Books, Totawa, NJ:, 1981.
- [17] Bertrand Russell. *Principles of Mathematics*. W.W. Norton and Company, inc, New York, 1903.
- [18] S. Smith. Are instantaneous velocities real and really instantaneous? *Studies in History and Philosophy of Modern Physics*, 34:261–280, 2003.
- [19] M. Tooley. In defense of the existence of states of motion. *Philosophical Topics*, 16(1):225–254, 1988.

# NENHUMA DIVERGÊNCIA METAFÍSICA SEM INCOMPATIBILIDADE LÓGICA<sup>1</sup>

Daniel Durante<sup>2</sup>

## Abstract

O objetivo deste artigo é defender a incompatibilidade lógica das posições divergentes como critério para caracterização dos desacordos genuinamente metafísicos. Ou seja, pretendo defender que uma disputa específica é um desacordo metafísico legítimo apenas quando as posições adversárias são regidas por lógicas distintas. Se correto, este critério poderia não apenas ajudar a distinguir disputas meramente verbais de debates metafísicos genuínos, como também forneceria um argumento contra o deflacionismo da metafísica, garantindo a sua substancialidade e relevância. Pretendo esclarecer o critério, apresentar seus fundamentos e compromissos filosóficos mais básicos, dar algumas motivações tanto lógicas quanto metafísicas para sua adoção, além de alguns exemplos de sua aplicação.

## 1. Introdução

Costumamos classificar como metafísicas questões cujas respostas envolvem uma caracterização da estrutura mais geral da realidade. Os objetos materiais, por exemplo, existem completamente em cada instante do tempo, ou existem apenas parcialmente a cada instante, tendo partes temporais? Os números existem independentemente de nosso pensamento ou são constituídos por nossa atividade mental? Estas questões costumam ser consideradas metafísicas porque suas respostas exigem esclarecimentos sobre como a realidade é estruturada e o que a compõe. Respostas à primeira exigem uma concepção sobre o que é o tempo e de que modo a sua passagem afeta e participa da estrutura da realidade, e respostas à segunda exigem uma concepção sobre se a estrutura da realidade tem ou não lugar para coisas que entendemos com o pensamento, mas que não têm materialidade, tais como os números.

Por serem tão gerais, as perguntas metafísicas como estas nos deixam, muitas vezes, com a sensação de que qualquer resposta que dermos a elas não terá efeito algum para fora dos limites da própria discussão. Afinal, dois mais dois continuará sendo quatro, sejam os números construções mentais ou objetos abstratos que existem fora de nossas mentes. O computador no qual escrevo estas palavras agora não se altera, não fica mais rápido ou mais lento, melhor ou pior, quando considero que ele está aqui a minha frente em toda sua existência, ou quando considero que o que tenho a minha frente é apenas uma pequena parte temporal do computador, cuja existência completa inclui suas partes passadas e futuras.

Estas considerações nos levam a perguntar: haverá alguma substancialidade nos debates metafísicos ou todos eles são mera perda de tempo sem qualquer conexão com o mundo do lado de fora dos departamentos de filosofia? Será que nossas concepções sobre a estrutura mais geral da realidade têm alguma consequência que extrapola os limites das

---

<sup>1</sup> An English version of this article is available at: <http://tiny.cc/b4y47y>.

<sup>2</sup> Departamento de Filosofia, UFRN. Email: [durante10@gmail.com](mailto:durante10@gmail.com).

próprias discussões? Se eu e você divergirmos sobre o papel da mente na consituição da realidade matemática, será que esta nossa divergência terá alguma influência na própria matemática ou em suas aplicações ou nos demais campos de nossas vidas?

A maioria das pessoas comuns, dos cientistas e até alguns filósofos consideram que a alternativa da irrelevância e da perda de tempo é a correta. Mas mesmo os que apreciam a metafísica, muitas vezes têm dúvidas sobre se uma suposta disputa específica é de fato relevante e substantiva ou é apenas um desacordo vocabular. Um buraco, por exemplo, tem independência ontológica, é algo que existe por si, ou sua existência é dependente de outra coisa, sendo ele apenas um aspecto da forma do material que o abriga? Ainda que eu seja alguém que considera as questões metafísicas substantivas e relevantes, eu não sou por causa disso obrigado a achar que todas as discussões supostamente metafísicas são de fato metafísicas. Talvez este assunto sobre buracos não seja uma questão metafísica genuína. Talvez ele seja um mero desacordo vocabular e não um problema que mereça o rótulo de problema metafísico.

E ainda que eu considere a metafísica relevante e também considere relevante e não meramente vocabular uma determinada questão específica, como saber se esta questão pertence mesmo ao campo teórico da metafísica? A definição que apresentei no início, de que as questões metafísicas são aquelas cujas respostas envolvem uma caracterização da estrutura mais geral da realidade, é muito ampla e vaga para funcionar como um critério. Haveria algum modo mais preciso de caracterizar as questões e debates metafísicos?

Façamos uma reorganização das questões até agora lançadas. A primeira e principal delas é a seguinte:

- (1) *A metafísica como disciplina tem alguma substancialidade ou as suas questões são irrelevantes para todos os assuntos fora de seus limites teóricos?*

Um modo bastante efetivo de responder à questão sobre a relevância e substancialidade de qualquer disciplina é analisar se as divergências teóricas internas à disciplina têm consequências externas, fora de seus limites teóricos. Se uma divergência teórica provocar outras divergências externas ao contexto teórico em que ela ocorreu, então esta divergência não é irrelevante. Não é um mero desacordo vocabular inofensivo, mas uma divergência substantiva que se espalha para outros assuntos. E se uma divergência teórica afeta outros assuntos, então a disciplina, a área de estudos em que esta divergência ocorre não será ela própria uma disciplina irrelevante.

Suponha, por exemplo, que eu e você tenhamos diferentes concepções filosóficas sobre o conceito de indivíduo humano. Suponha que divirjamos sobre o que caracteriza um indivíduo humano. Suponha, ainda, que de acordo com o meu modo de entender o assunto, um embrião de poucas semanas não seja ainda classificado como um indivíduo humano. O embrião não atende todos os critérios que minha concepção do que seja um indivíduo humano requer. E suponha que de acordo com o seu modo de entender, o mesmo embrião atende os seus critérios e é, por isso, caracterizado por você como um indivíduo humano. Então esta nossa divergência filosófica sobre como caracterizar um indivíduo humano nos leva a divergir sobre se um embrião de poucas semanas é ou não um indivíduo humano. Do mesmo modo, uma divergência que pudéssemos ter sobre o modo como as coisas materiais existem através do tempo poderia nos levar a divergir sobre se o computador em que digito estas palavras está completamente em minha frente, agora, ou se estou diante apenas de uma pequena parte temporal dele.

Uma consequência de nossa divergência sobre o que caracteriza um indivíduo humano é que se de acordo com o meu critério um embrião de poucas semanas não é um indivíduo

humano, então eu tenho condições de defender a moralidade e legalidade do aborto sem contradizer o princípio geral de que todo o indivíduo humano tem direito à vida. Já você, ao considerar o embrião como um indivíduo humano, não tem como defender a moralidade e legalidade do aborto sem contradizer o mesmo princípio. Nossa divergência sobre o que caracteriza um indivíduo humano provoca, então, outras divergências exteriores ao seu contexto teórico, já que por causa dela vamos divergir também sobre a moralidade e a legalidade do aborto. Por causa dela divergiremos sobre ética, direito, políticas de saúde pública, e até nossas atitudes pessoais diante de situações específicas que envolvam o aborto ou sua possibilidade também divergirão. Este caso mostra que a disciplina filosófica que abriga nossa divergência sobre como entender o conceito de indivíduo humano é bastante relevante e substancial.

Do mesmo modo, com relação à nossa divergência metafísica sobre a relação entre a existência e a duração temporal, que nos levou a divergir sobre se o que tenho a minha frente é o computador em toda sua existência ou apenas uma parte temporal dele; se mostrarmos que esta divergência tem consequências exteriores à metafísica, se esta divergência levar a outras divergências na física, na engenharia, em outros assuntos, se ela se espalhar, então saberemos que a metafísica é relevante e substancial, porque teremos mostrado que divergências metafísicas têm consequências que extrapolam os seus próprios limites teóricos.

Deste modo, daremos uma resposta à questão (1) acima que garante substancialidade e relevância à metafísica se respondermos afirmativamente à seguinte questão:

- (2) *As divergências metafísicas têm consequências que extrapolam os limites teóricos da própria filosofia?*

A metafísica não será uma perda de tempo irrelevante se mostrarmos que nossas divergências sobre questões metafísicas espalham-se e têm consequências em outras áreas externas ao âmbito das teóricas discussões filosóficas. Mas para que uma resposta à questão (2) possa fornecer uma resposta adequada à questão (1) precisamos, primeiramente, de um critério para decidir quando uma divergência específica deve ou não ser considerada como sendo de natureza metafísica. Antes de responder à questão (2) precisamos, então, responder à seguinte questão:

- (3) *O que caracteriza uma proposta específica como pertencente à metafísica e uma divergência específica como divergência metafísica?*

Minha proposta é utilizar o critério que intitula este artigo, *nenhuma divergência metafísica sem incompatibilidade lógica*, como uma caracterização das propostas e divergências metafísicas que responde à questão (3) de um modo que leva a uma resposta afirmativa da questão (2) e, por conseguinte, a uma resposta da questão (1) que assegura relevância e substancialidade à metafísica. Ao raciocinarmos de acordo com lógicas diferentes e incompatíveis, usamos diferentes padrões para justificar a verdade das sentenças. Se  $P_1, \dots, P_n \vdash A$  for um argumento válido de acordo com a lógica que eu uso, mas inválido na lógica que você usa, e se nós dois concordamos que todas as premissas  $P_1, \dots, P_n$  são verdadeiras, ainda assim nós poderemos discordar sobre a verdade de  $A$ . Para mim  $A$  tem que ser verdadeira, pois em minha lógica  $P_1, \dots, P_n$  justificam sua verdade. Mas, para você,  $A$  pode ser falsa, porque na sua lógica  $P_1, \dots, P_n$  não são suficientes para justificar a verdade de  $A$ . Nossa divergência lógica nos autoriza a discordar sobre o valor de verdade de  $A$ , que pode ser uma sentença de qualquer assunto. Então, ao demandarem desacordos

lógicos, os desacordos metafísicos podem espalhar-se através da lógica e ter consequências em todos os demais assuntos sobre os quais raciocinamos. Os desacordos metafísicos serão, portanto, substanciais e relevantes.

O fundamento deste critério é a ideia de que os princípios lógicos de inferência constituem-se em (ou são contrapartidas de) princípios metafísicos, e que qualquer proposta metafísica específica é indissociável de uma proposta lógica específica, de modo que não é possível haver divergência metafísica legítima sem que haja também divergência ou incompatibilidade lógica. Esclarecer este critério e seu fundamento, e apresentar algumas justificativas lógicas e metafísicas para sua adoção, além de alguns exemplos de sua aplicação, são o objetivo deste artigo.

A proposta que aqui apresento possui forte inspiração em dois importantes filósofos do Século XX. Willard van Orman Quine e Michael Anthony Eardley Dummett. Esta inspiração é ao mesmo tempo positiva e crítica. Por um lado, aproprio-me de muitas ideias destes autores e, por outro lado, ao reconhecer os problemas que a literatura tem apontado nas abordagens de cada um deles, construo o critério como uma resposta ou reação a eles.

No restante do artigo, na Seção 2, antes de propriamente esclarecer e justificar o critério da incompatibilidade lógica, reflito sobre seus limites e sobre a dupla posição que ele ocupa enquanto uma tese em filosofia da lógica e uma tese metametafísica. Na Seção 3, apresento brevemente a proposta metodológica de Quine para a ontologia. Suas ideias de arregimentação e compromisso ontológico representam a primeira inspiração de nossa tese e constituem-se em seus primeiros passos. Na Seção 4, esclareço uma limitação do método quineano e inicio a explicação mais minuciosa do critério, que corresponderá a uma proposta de superação desta limitação. Na Seção 5, discuto a proposta de Dummett para o tratamento das diversas instâncias do debate entre realismo e antirrealismo, o que fornecerá tanto um exemplo de aplicação bem sucedida do critério da incompatibilidade lógica, quanto algumas motivações metafísicas para sua adoção. A abordagem de Dummett aos debates metafísicos representa a segunda fonte principal de inspiração para nossa abordagem. Na Seção 6, critico a tese tradicional da separação entre lógica e metafísica, e aprofundo os argumentos em favor da interpretação dos teoremas quantificacionais da lógica como princípios metafísicos. Na Seção 7, proponho a tese da generalidade absoluta como um critério de demarcação da lógica e reforço, com argumentos lógicos, a concepção dos princípios lógicos como princípios metafísicos. Na Seção 8, aplico o critério a uma controvérsia específica entre Kris McDaniel e Peter van Inwagen a respeito da univocidade ou não do ser, demonstrando que, neste caso, a controvérsia não se trata de uma divergência metafísica legítima. Finalmente, na Seção 9, termino o artigo com algumas breves considerações finais.

### **2. Alcance e Limites: filosofia da lógica e metametafísica**

A partir da primeira metade do Século XX, a ideia de uma suposta neutralidade metafísica da lógica ganhou força e adeptos. Ela já estava presente nas *Investigações Lógicas* de Edmund Husserl (2015) e adquiriu notoriedade com as teses do *Tractatus Logico-Philosophicus* de Ludwig Wittgenstein (2014) de que “o mundo é a totalidade dos fatos, não das coisas”, e de que as verdades lógicas nada informam sobre os fatos, não têm conteúdo factual. Posteriormente, Rudolf Carnap (1969) deu forma definitiva a esta ideia e a utilizou para isolar e proteger a lógica das pesadas críticas deflacionistas que ele e os demais positivistas lógicos dirigiram à metafísica. Finalmente, com a adoção da lógica

clássica de primeira ordem e das concepções de Alfred Tarski (1944) sobre verdade e consequência lógica como o padrão ortodoxo entre lógicos e filósofos, a tese da neutralidade metafísica da lógica tornou-se hegemônica e ubíqua.

Para ilustrar esta tese, costumo contar aos meus estudantes a fábula do meteorologista de um jornal local que, no primeiro dia de trabalho, não queria arriscar seu emprego com uma previsão do tempo errada e publicou em seu boletim: “amanhã, em Natal, choverá ou não choverá”. Ele acertou a previsão, mas perdeu o emprego mesmo assim. A previsão “choverá ou não choverá” é tão geral que simplesmente é impossível de ser falsa. Chova ou faça sol ela está correta. Mas qual é a utilidade de uma previsão do tempo que, ainda que verdadeira, não nos ajuda a decidir se ao sair de casa pela manhã devemos ou não levar o guarda-chuva ou um casaco? Ao ganhar generalidade a previsão perdeu informação, e o pobre meteorologista, o emprego.

Por serem compatíveis com todas as possibilidades em que os fatos podem se dar, as verdades lógicas, tais como “choverá ou não choverá”, estariam separadas deles e da realidade e seriam, por isso, verdades meramente analíticas. Elas não tocam os fatos e não nos ajudam a saber o que ocorrerá ou não ocorrerá. Ao nada nos informarem sobre a realidade, as verdades lógicas seriam metafisicamente neutras.

Apesar da predominância desta tese durante boa parte do Século XX, os principais filósofos ao longo da história anterior, de Aristóteles a Russell, passando pelos medievais, por Leibniz e por Frege, não concebiam a lógica como metafisicamente neutra. Para a surpresa de muitos lógicos contemporâneos os princípios da identidade, da contradição e do terceiro excluído, que contemporaneamente são conhecidos como os três princípios aristotélicos da lógica clássica (D’Ottaviano and Feitosa, 2003), não foram apresentados por Aristóteles como princípios lógicos ou princípios da razão. Foram apresentados como “os princípios mais certos aplicáveis a todas as coisas” (Metafísica, 1005b10-35) (Aristotle and McKeon, 1941). Ou seja, princípios metafísicos. Um indício deste fato é que os referidos princípios são propostos e detalhadamente tratados não nos textos lógicos de Aristóteles, agrupados pelos comentadores antigos no *Organon*, mas no Livro 4 da *Metafísica*. Lá eles aparecem como primeiros princípios da ciência do ser enquanto ser (metafísica) e, por isso, de todas as outras ciências (Smith, 2019).

Em Frege também podemos encontrar vestígios desta concepção metafísica da lógica. Ele considerava que só seria possível justificar a concepção de que as leis da lógica são leis prescritivas para o pensamento, se as entendêssemos de um modo mais fundamental como leis da verdade, ou leis que estabelecem o que é. Para explicar-se distinguiu dois sentidos para a palavra “lei”. Num sentido, o das leis da natureza, uma lei estabelece o que é. Todas as ocorrências factuais sempre se conformam às leis da natureza. No outro sentido, o das leis morais ou civis, uma lei é uma prescrição que estabelece o que deve ser. Nem todas as ocorrências factuais se conformam com as leis morais ou civis. Como o pensamento é um processo mental, as leis sobre como o pensamento é estão no âmbito da psicologia, não da lógica. Então, quando se diz que as leis da lógica são leis do pensamento, a única possibilidade é de as entendermos como leis prescritivas para o pensamento.

Mas o que dá à lógica o poder de prescrever as regras do pensamento? A resposta de Frege foi identificar as leis da lógica com as leis da verdade, leis que no âmbito mais geral possível estabelecem o que é.

Qualquer lei que estabelece o que é pode ser concebida como prescrevendo que se deve pensar em conformidade com ela, e é, então, uma lei [prescritiva] do pensamento. (Frege and Furth, 1967, 12)

Assim, o que dá à lógica o direito de prescrever o pensamento correto é, para Frege, o fato de as leis lógicas serem leis sobre o que é. Deste modo, para Frege, assim como para Aristóteles, as leis da lógica são leis sobre todas as coisas, sobre o que é. E, neste sentido, são leis metafísicas.

Nos dias atuais, com a explosão das propostas de lógicas não clássicas, a tese da neutralidade metafísica da lógica, apesar de ainda vigente e forte em muitos círculos, vem sendo regularmente e cada vez mais confrontada.<sup>3</sup> Podemos talvez considerar que a sobrevivência ainda hoje da tese da neutralidade metafísica da lógica é uma das últimas e mais fortes influências do positivismo lógico na filosofia contemporânea.

O critério que pretendo apresentar, motivar e exemplificar e que intitula este artigo assume exatamente a antítese da neutralidade metafísica da lógica. Mais do que a sua não neutralidade metafísica, quero defender a ideia oposta de que os princípios lógicos de inferência são também princípios metafísicos, e que qualquer proposta metafísica específica é indissociável de uma proposta lógica específica, não sendo possível haver uma divergência metafísica sem a companhia de uma divergência lógica.

Antes de esclarecer e exemplificar seu conteúdo, vamos refletir um pouco sobre como entender a dupla natureza filosófica do critério que proponho. Ele inclui tanto uma tese em filosofia da lógica, quanto uma tese metafísica, ou *metametafísica*, como alguns filósofos preferem chamar as teses que lidam com a natureza e a metodologia das questões metafísicas (Tahko, 2015). No âmbito da filosofia da lógica, a tese que nosso critério carrega é bastante simples. Ela alia-se à tradição de Aristóteles e Frege e constitui-se na antítese da tese da neutralidade metafísica da lógica. Os princípios lógicos de inferência não são metafisicamente neutros. Ao contrário, eles são princípios metafísicos. Ao *prescreverem* as regras do pensamento, os princípios lógicos estabelecem obrigações e proibições vigentes a todos os seres e podem ser interpretados como princípios metafísicos que *descrevem* a estrutura mais geral da realidade.

Em seu aspecto metametafísico, por sua vez, nosso critério apresenta respostas a duas questões fundamentais. Ele tanto (*a*) responde afirmativamente à questão da substancialidade e relevância da metafísica, quanto propõe que (*b*) os métodos da lógica podem ser proveitosamente incluídos entre os métodos da investigação metafísica.

Quanto à questão (*a*), se o que caracteriza um debate como genuinamente metafísico é a incompatibilidade lógica das posições adversárias, então todos os desacordos metafísicos são substanciais e fazem diferença. Afinal, se diferentes propostas metafísicas demandam lógicas diferentes, então nossos posicionamentos metafísicos, através de suas especificidades lógicas, influenciarão todos os demais assuntos sobre os quais raciocinamos. Os diferentes modos de fazer inferência que a divergência metafísica requer fazem com que esta divergência seja carregada para fora do limites da metafísica. Ela influenciará todos os assuntos, dos mais corriqueiros aos mais sofisticados. Então, se eventualmente nos deparamos na literatura com debates supostamente metafísicos, mas que não têm consequências que extrapolam seus próprios limites, isso só pode ter ocorrido quando as posições adversárias não são logicamente incompatíveis. Mas neste caso, segundo nosso critério, este não será um genuíno debate metafísico. Certamente a literatura está repleta de exemplos deste caso, e uma das principais motivações de nossa proposta é justamente separar estas discussões meramente verbais dos debates genuinamente metafísicos.

Nosso critério também responde à questão (*b*), sobre os métodos da investigação metafísica, propiciando uma alternativa metodológica específica. E aqui é preciso ter bastante

---

<sup>3</sup> Um excelente exemplo desta tendência é Williamson (2015).

cuidado. Por um lado, o critério é suficientemente atrevido para afirmar que qualquer proposta que mereça o rótulo de “metafísica” deve ter uma especificidade lógica, de modo que a toda proposta metafísica específica relaciona-se, como contrapartida, uma proposta lógica específica. Isso sem dúvida dá à lógica um papel metodológico privilegiado na investigação metafísica. Mas este papel privilegiado não deve ser confundido com exclusividade. De maneira nenhuma estou propondo o fim da metafísica e sua substituição pela lógica, nem tampouco pretendo legislar sobre os métodos legítimos nos quais a investigação metafísica deve ocorrer. As propostas metafísicas são produzidas e argumentadas pelos filósofos de diversas maneiras e em diversos estilos e tradições. E suas propostas e argumentações podem ser fortes e convincentes o suficiente para angariar adeptos, suplantando posições rivais e solucionar questões, sem que haja qualquer necessidade de apelo explícito à lógica. Esta saudável liberdade metodológica, no entanto, não impede ninguém de aplicar nosso critério para acessar, interpretar e fornecer a contrapartida lógica de uma proposta metafísica qualquer e, ao fazer isso, obter esclarecimento e compreensão suficientes para julgar, avaliar, decidir, completar e até apontar falhas na proposta original, nos termos em que ela foi originalmente produzida. Esta mesma liberdade metodológica também não impede ninguém de procurar na miríade atual de sistemas lógicos específicos, produzidos com as mais variadas motivações, que na maioria das vezes não têm a mais remota ligação com a metafísica, quais seriam as interpretações metafísicas específicas a eles vinculadas, retirando assim dos desenvolvimentos da lógica, argumentos e conteúdo metafísico.

Podemos, a título de ilustração, comparar a relação entre lógica e metafísica suscitada por nosso critério com o isomorfismo que Descartes mostrou haver entre álgebra e geometria, quando introduziu o sistema de coordenadas cartesianas, lançando as bases da geometria analítica. O surgimento da geometria analítica não destruiu nenhuma das duas disciplinas em que a matemática se dividia. Depois de Descartes, a álgebra continuou sendo álgebra e continuou sendo desenvolvida independentemente da geometria, tendo as mais variadas motivações e aplicações. Similarmente, a geometria continuou sendo geometria e continuou sendo desenvolvida independentemente da álgebra. No entanto, a geometria analítica possibilitou o uso da álgebra para a solução de problemas geométricos e o uso da geometria para a solução de problemas algébricos. A geometria analítica nos assegura que todo fato geométrico tem uma contraparte algébrica e vice-versa. Ela inclusive nos permitiu ampliar os horizontes da geometria para além das três dimensões de nossa intuição espacial, o que teve profundas e profícuas consequências em suas aplicações na física, por exemplo.

O critério que aqui defendo funda-se em uma hipótese que pode ser entendida, metaforicamente, como a admissão de um isomorfismo entre lógica e metafísica nos mesmos termos do isomorfismo entre álgebra e geometria que o sistema de coordenadas cartesianas nos dá. A lógica, então, seria para a metafísica o mesmo que a álgebra é para a geometria. O desenvolvimento da lógica é guiado pela pauta que os lógicos dão à sua disciplina, do mesmo modo que o desenvolvimento da álgebra é guiado pela pauta que os algebristas dão para a sua disciplina. Entre as muitas motivações e aplicações que pautam o trabalho dos algebristas hoje, há também as geométricas. Do mesmo modo, se a hipótese do isomorfismo entre lógica e metafísica for correta, então deveremos encontrar aplicações metafísicas nos desenvolvimentos já feitos em lógica, e poderemos incluir na pauta dos futuros desenvolvimentos em lógica também motivações e aplicações metafísicas.<sup>4</sup>

---

<sup>4</sup> No decorrer do texto usarei a expressão “critério da incompatibilidade lógica” para me referir ao critério de caracterização das divergências genuinamente metafísicas que intitula este artigo, e utilizarei



Não estou de modo algum propondo o fim da metafísica e sua conversão em lógica. Nem tampouco estou propondo que todas as justificativas e motivações para as pesquisas em lógica sejam restritas à metafísica. Mas estou sim propondo que podemos olhar para as propostas metafísicas procurando suas contrapartidas lógicas, e que ao encontrá-las obteremos elementos que eventualmente podem melhorar nosso entendimento sobre estas propostas e podem nos ajudar a distinguir controvérsias genuinamente metafísicas de divergências meramente verbais ou de outra natureza.<sup>5</sup>

Dada a imensa amplitude e generalidade das propostas e disputas metafísicas e dos desenvolvimentos que hoje em dia são considerados fazer parte da lógica, estou ciente das enormes dificuldades que o critério que proponho enfrenta. Ainda não tenho respostas para a maioria delas. Também não espero angariar a simpatia dos metafísicos avessos a propostas fundadas em arregimentação formal, nem daqueles envolvidos em debates que segundo nosso critério não seriam classificados como genuinamente metafísicos. Aos primeiros, reforço que não proponho a substituição da metafísica pela lógica, mas apenas o uso de um critério lógico para a identificação da substancialidade dos debates metafísicos. Este critério, longe de substituir a metafísica e seus métodos, de atacá-la, reformá-la ou decretar seu fim, é, ao contrário, uma defesa de sua centralidade e importância. Ao vincularem-se a posicionamentos lógicos específicos, as proposições metafísicas garantem sua relevância e influência para fora de seus próprios limites, atingindo todos os demais assuntos sobre os quais raciocinamos. Aos segundos, aqueles filósofos envolvidos em debates que não seriam genuinamente metafísicos de acordo com nosso critério, acrescento que não ser classificado como metafísico não é, em geral, um demérito. Para a maioria das pessoas, na maioria das vezes, é exatamente o contrário disso, um mérito. O debate pode ser científico, religioso, cultural e, nestes casos, será quase sempre relevante. E ainda que o debate seja meramente verbal ou vocabular, pode ainda assim ter relevância ao esclarecer determinados usos e abusos das palavras. Talvez a própria tese que aqui defendo seja, no final das contas, ela mesma uma proposta meramente vocabular sobre o que deve ou não ser incluído sob o conceito de metafísica. Ainda que este seja o caso, a proposta não seria irrelevante e sem mérito apenas por isso.

Também estou ciente de que o critério que proponho envolve escolhas metodológicas e temáticas sobre como tratar cada uma das duas disciplinas, que certamente desagradarão a muitos. Lógicos e metafísicos poderão alegar que tanto a lógica quanto a metafísica são muito diversas ou mais amplas do que aquilo que estou incluindo sob a hipótese do isomorfismo. Meu único consolo sobre este ponto é negativo. As questões de delimitação são extremamente controversas e não há, que eu saiba, nenhuma proposta de delimitação destas duas disciplinas que tenha angariado adesão hegemônica. O que no momento acredito ter condições de oferecer é nada mais do que uma definição clara do critério, alguns esclarecimentos de seu significado, de seus compromissos e pressuposições filosóficas, juntamente com um conjunto de motivações tanto lógicas quanto metafísicas para a sua adoção, além de uns poucos exemplos de sua aplicação.

---

a expressão “hipótese do isomorfismo” para me referir a esta relação isomórfica entre lógica e metafísica que proponho como fundamento para o critério da incompatibilidade lógica.

<sup>5</sup> Uma outra consequência da admissão da hipótese do isomorfismo é que ela nos incita a fazer também a pesquisa inversa desta que ora nos ocupa. Ou seja, podemos olhar para as propostas lógicas procurando suas contrapartidas metafísicas e, se as encontrarmos, obteremos elementos que eventualmente podem melhorar nosso entendimento sobre estas propostas, ajudando-nos a esclarecê-las, justificá-las e classificá-las. O critério oposto ao nosso título, *nenhuma divergência lógica sem incompatibilidade metafísica*, também merece nossa atenção e pesquisa. Mas isto é assunto para um outro artigo.

### 3. Metaontologia Quineana: compromisso ontológico

Afirmar que o que caracteriza uma proposta como metafísica é a sua especificidade lógica, que não há divergência metafísica sem divergência lógica, evidencia um forte compromisso com a formalização e vincula nossa abordagem a uma tradição que teve na proposta metodológica de Quine para a ontologia seu modelo paradigmático. Mas o que é ontologia? Para a maioria dos filósofos a ontologia é uma subárea da metafísica. A metafísica lida com os aspectos da estrutura mais geral da realidade. Quais coisas ou tipos de coisas existem é um destes aspectos, e é o tema da ontologia.<sup>6</sup> Nosso critério vincula-se à proposta metodológica de Quine e serve-se de muitos dos seus elementos. Os próximos passos serão, por isso, entender as linhas gerais da proposta de Quine, o que fazemos de modo bastante abreviado nesta sessão, e entender alguns de seus limites e problemas, o que fazemos na sessão seguinte.<sup>7</sup>

Aquilo que tomamos como verdadeiro nos comprometemos. Quine aproveitou esta platitude para propor uma metodologia naturalista para a ontologia. Ele introduziu um critério para identificar quais são os *compromissos ontológicos* das teorias que aceitamos e propôs que nossa ontologia deve conter exatamente as entidades que correspondem a estes compromissos ontológicos extraídos de nossas melhores teorias científicas.

Segundo Quine, ao assumirmos uma teoria como verdadeira, nos comprometemos com a existência de certas entidades que são os compromissos ontológicos da teoria:

Estamos condenados a uma pressuposição ontológica particular se, e somente se, o alegado pressuposto tiver que ser reconhecido entre as entidades sobre as quais variam nossas variáveis a fim de tornar uma de nossas afirmações verdadeiras.

Quine (1963c, 13)

Ou seja, uma teoria compromete-se ontologicamente com espécies biológicas, por exemplo, quando a não inclusão de espécies biológicas entre as entidades que são valores das variáveis de suas sentenças faz com que algumas destas sentenças fiquem falsas.

Quine, seguindo a tradição inaugurada por Frege e Russell, vincula existência com quantificação. Esta vinculação faz com que estas entidades (que têm que estar entre os valores das variáveis das sentenças de nossas teorias para que elas sejam verdadeiras) sejam reveladas nas afirmações existenciais. Ele afirma explicitamente, “A existência é o que o quantificador existencial expressa. Há coisas do tipo  $F$  se e somente se  $\exists x F(x)$ ” (Quine, 1969, 97). Então uma teoria compromete-se ontologicamente com espécies biológicas se a afirmação da existência de espécies biológicas estiver entre as sentenças da teoria. Em geral, uma teoria  $\mathbf{T}$  compromete-se ontologicamente com entidades do tipo  $P$  se e somente se a afirmação de que existem  $P$ s está entre as afirmações da teoria. Em símbolos:

$$\mathbf{T} \text{ compromete-se ontologicamente com } Ps \iff \mathbf{T} \models \exists x P(x)$$

Outro fato importante à proposta de Quine, muitas vezes negligenciado, é que os portadores de compromissos ontológicos não são as sentenças tomadas isoladamente, mas *teorias*. E por teoria Quine entende sua definição lógica usual. Uma teoria é um conjunto de sentenças fechado pela relação de consequência lógica. Ou seja, dado um conjunto de sentenças  $\mathbf{T}$  qualquer, a teoria  $\mathbf{T}$  (da qual  $\mathbf{T}$  é uma axiomática) contera além das sentenças

<sup>6</sup> A distinção e delimitação entre ontologia e metafísica é também um assunto polêmico. Estou provisoriamente assumindo de modo acrítico a posição hegemônica entre os filósofos. A ontologia é uma subárea da metafísica que responde à questão sobre quais coisas existem. Nas próximas seções teremos mais a dizer sobre a distinção entre metafísica e ontologia.

<sup>7</sup> Em Berto and Plebani (2015), van Inwagen (2009) e Durante (2014) há detalhadas e acessíveis exposições dos diversos elementos que compõem a proposta metodológica de Quine.

de  $T$  todas as suas consequências lógicas.<sup>8</sup> Então tratar um determinado discurso como uma teoria, ou seja, acrescentar ao discurso todas as suas consequências lógicas, nos dá a capacidade de identificar nas afirmações existenciais desta teoria todos os compromissos ontológicos do discurso.

Mas este é um critério lógico e formal. Nem a linguagem natural e nem mesmo as linguagens científicas semiformalizadas contém explicitamente variáveis e quantificadores existenciais. Então, antes de encontrar os compromissos ontológicos de uma teoria, é preciso arregimentá-la em uma notação canônica formal.

O método da ontologia, para Quine, requer assim três etapas:

- (1) Arregimente as melhores teorias científicas em uma notação canônica formal.
- (2) Use o critério definido acima para listar todos os compromissos ontológicos (dados pelas afirmações existenciais) destas teorias arregimentadas.
- (3) Inclua em sua ontologia apenas as entidades correspondentes a estes compromissos ontológicos e nada mais.

A proposta de Quine para a ontologia é naturalista. A filosofia trabalha junto com a ciência. A ontologia, aliás, vem depois da ciência. Primeiro me dê as nossas melhores teorias científicas, depois te dou minha ontologia, dirá a filósofa (ou filósofo) engajada à proposta metodológica quineana.

Esta filósofa não tem qualquer liberdade sobre os passos 2 e 3. Eles são fixos. Correspondem a meras aplicações de procedimentos preestabelecidos. O espaço para o debate filosófico sobre ontologia ocorre, então, exclusivamente no passo 1 da estratégia, o passo da arregimentação através de paráfrases na notação canônica. Diferentes paráfrases podem levar a diferentes compromissos ontológicos. Mas como, segundo Quine, devemos sempre respeitar a navalha de Occam, então nossa ontologia deveria ser sempre aquela que se compromete com menos entidades. Incluiremos em nossa ontologia apenas aquelas entidades que forem *indispensáveis*.

Um debate ontológico sobre se determinado tipo de entidade  $P$  existe, seria sempre um debate sobre a indispensabilidade  $P$ . Seja  $T$  a teoria científica que lida com  $P$ s. E sejam  $T_1, \dots, T_n$  todas as arregimentações formais disponíveis de  $T$ . Se todas as arregimentações têm entre suas consequências lógicas a afirmação existencial  $\exists x P(x)$ , ou seja, se para todo  $(1 \leq i \leq n)$ ,  $T_i \models \exists x P(x)$ , então as entidades  $P$  são indispensáveis e devem entrar em nossa ontologia. Caso haja alguma arregimentação  $T_j$ ,  $(1 \leq j \leq n)$  em que  $\exists x P(x)$  não é uma de suas consequências lógicas, ou seja, caso haja algum  $j$  tal que  $T_j \not\models \exists x P(x)$ , então as entidades do tipo  $P$  não são indispensáveis e não devem entrar em nossa ontologia.

O debate ontológico resume-se, na metodologia quineana, às discussões dos filósofos sobre se diferentes propostas de arregimentação são ou não bem-sucedidas em parafrasear todo o conteúdo das teorias científicas. Dadas duas arregimentações bem-sucedidas neste quesito, ganha o debate ontológico aquela que assumir menos compromissos, a que for mais econômica ontologicamente.

Por exemplo, se for possível apresentar arregimentações formais de nossas melhores teorias científicas relacionadas com espécies biológicas que não possuam entre suas consequências lógicas a afirmação da existência de espécies biológicas, isto demonstrará que

<sup>8</sup> Em símbolos:  $\mathbf{T} = \{\varphi / T \models \varphi\}$ . Uso a notação  $T$ , grafada sem serifa, para denotar um conjunto qualquer de sentenças (um discurso), e a notação  $\mathbf{T}$ , em **negrito**, para denotar o fecho dedutivo de  $T$ . Ou seja,  $\mathbf{T}$  denota a teoria completa da qual as sentenças de  $T$  são os axiomas. Deste modo a definição de compromisso ontológico apresentada acima também poderia ser dada por:

$$\mathbf{T} \text{ compromete-se ontologicamente com } Ps \iff \exists x P(x) \in \mathbf{T}$$

a biologia não se compromete ontologicamente com espécies biológicas e elas não devem fazer parte de nossa ontologia, porque não são indispensáveis. Para Quine nossa ontologia deve conter apenas aquelas entidades que são indispensáveis às nossas melhores teorias científicas.

Já os números, por sua vez, Quine os incluía na ontologia, porque os via como indispensáveis. Muitas de nossas melhores teorias científicas comprometem-se ontologicamente com eles ao exigirem que números estejam entre os valores das variáveis quantificadas em suas afirmações existenciais. Os números deveriam, por isso, fazer parte da realidade e estar entre as coisas que existem.

Mas para que um debate ontológico deste tipo possa ser racionalmente conduzido, é preciso que seus participantes concordem previamente sobre algumas importantes questões fundamentais:

- a) Os filósofos participantes do debate têm que concordar sobre quais são as nossas melhores teorias científicas, aquelas que merecem a consideração da filosofia.
- b) Eles também têm que concordar sobre qual é a linguagem formal em que as arregimentações são feitas. Quais são os recursos formais que podem ser utilizados em nossas paráfrases arregimentadas?
- c) Por fim, eles têm que concordar sobre qual é a lógica que governa todas as teorias. Afinal, lógicas diferentes apontarão diferentes sentenças entre as consequências de determinada teoria, o que pode levar a compromissos ontológicos diferentes.

As prescrições da metodologia quineana para estes acordos são bastante específicas e restritivas. Segundo ele, o item (a) é de responsabilidade dos cientistas não dos filósofos. Na ontologia naturalizada quineana não há espaço para a especulação filosófica *a priori* sobre o que existe. O trabalho dos filósofos começa apenas após o acordo científico sobre quais são as nossas melhores teorias.

Com relação ao item (b) Quine estipula que a arregimentação deve ser feita na linguagem da lógica de primeira ordem, com identidade, sem constantes individuais, tratando os nomes via teoria das descrições definidas de Russell. Esta linguagem formal na qual as arregimentações de todas as teorias científicas devem ser feitas para que seus compromissos ontológicos possam ser revelados ficou conhecida como a “notação canônica” de Quine. Finalmente, quanto ao item (c), a única possibilidade legítima para Quine é a Lógica Clássica de Primeira Ordem, com o eventual complemento de sua teoria de conjuntos de primeira ordem NF.<sup>9</sup>

Esta proposta de Quine para a metodologia da ontologia tem muitos méritos. Sua noção geral de compromisso ontológico juntamente com os requerimentos da arregimentação formal e da indispensabilidade ajudaram a esclarecer inúmeras questões, tais como a suposta contradição das declarações de não-existência (Quine, 1963c). No entanto, suas prescrições para os acordos (a), (b) e (c) jamais ganharam adesão hegemônica.

No caso do acordo (a), ainda que haja grande consenso em reconhecer que os compromissos ontológicos das teorias científicas devem ser seriamente considerados em qualquer proposta ontológica, muitos filósofos com inclinação menos naturalista não estão dispostos a aceitar que a ciência tenha exclusividade em ditar a pauta da ontologia. Mais graves do que esta divergência de postura são as limitações que as prescrições de Quine para os acordos (b) e (c) impõem à sua abordagem. Na próxima sessão explorarei uma destas limitações específicas e, ao final, proporei um modo de transpô-la que nos levará

<sup>9</sup> Ver Quine (1963b) e Durante (2011).

diretamente ao esclarecimento do critério da incompatibilidade lógica e da hipótese do isomorfismo.

#### 4. Além da Ontologia: compromisso metafísico

Suponha que estejamos interessados em investigar se os números existem independentemente de nosso pensamento ou se são constituídos exclusivamente por nossa atividade mental. Esta questão pode ser entendida de duas maneiras distintas, ou como uma pergunta ontológica mais restrita sobre se os números existem ou não, se fazem ou não parte da realidade; ou como uma pergunta metafísica mais ampla sobre qual o papel da mente na constituição da realidade matemática.

Por exemplo, algumas pessoas dirão que se os números são construções que só existem em nosso pensamento, então eles não fazem parte da realidade objetiva, na verdade não existem. Para estas pessoas, o pensamento está completamente separado da realidade e, portanto, o que está só no pensamento não existe na realidade. Para elas, então, a divergência sobre se os números são construções mentais ou são independentes de nosso pensamento não será uma questão metafísica ampla sobre a estrutura da realidade, mas uma questão ontológica restrita sobre se os números existem ou não. Se eles forem construções mentais, não existem. Se forem independentes de nosso pensamento, existem.

Entendida deste modo ontológico, a questão das entidades matemáticas é tratável pela metodologia de Quine, que nos ajudará a respondê-la em um de dois modos possíveis: ou os números existem, porque eles estão entre as afirmações existenciais das arregimentações de nossas melhores teorias científicas e são, por isso, indispensáveis; ou eles não existem, porque obtivemos sucesso na produção de arregimentações alternativas para as teorias científicas que não assumem compromissos ontológicos com os números. Quine admitiu a indispensabilidade dos números e os incluiu em sua ontologia. Mas para ele, tanto quanto a ciência, a ontologia e toda a filosofia também são revogáveis. Se algum dia conseguirmos produzir arregimentações alternativas e bem-sucedidas de nossas melhores teorias científicas que não assumam compromissos ontológicos com números, neste dia deveremos retirar os números de nossa ontologia. Alguns filósofos adotaram esta pauta e se lançaram nesta e em outras tarefas semelhantes com resultados parciais bastante interessantes.<sup>10</sup>

Mas seja qual for a resposta aqui, esta interpretação ontológica da questão fundamenta-se em um pressuposto mais geral sobre a estrutura da realidade: o que existe apenas em nossa mente, aquilo que é constituído apenas por nosso pensamento, não faz parte da realidade, não existe. Então a interpretação ontológica da questão compromete-se com a seguinte tese metafísica a respeito do papel da mente na estrutura geral da realidade, que poderíamos chamar de *tese realista*:

(1) *A mente está separada da realidade.*

A outra interpretação possível para a questão da relação dos números com o pensamento não a trata como uma questão ontológica restrita sobre a existência ou não dos números, mas como uma questão metafísica ampla sobre se a mente tem ou não papel na constituição da realidade matemática. Algumas pessoas podem não aceitar a tese realista (1) e por isso considerar que ser constituído por nossa mente faz parte dos traços admissíveis para a estrutura da realidade. Ou seja, para estas pessoas uma das características da realidade é que a mente pode ter papel constitutivo em pelo menos alguma parte sua, sustentando a existência de algumas das coisas que consideramos reais. Para elas, então, afirmar que

---

<sup>10</sup> Um ótimo exemplo de trabalhos deste tipo é Field (2016).

os números são construções mentais não significa afirmar que eles não existem, significa ao invés afirmar que eles existem, e que a mente tem papel constitutivo em sua existência.

Mesmo que eu discorde desta posição e queira debatê-la, o debate que travarei não será ontológico. Não será um debate sobre se os números existem ou não, mas um debate metafísico sobre o papel da mente na estrutura da realidade. As duas respostas possíveis para esta interpretação da questão são: ou a nossa mente não tem papel algum na constituição da realidade matemática e os números, se existirem, existirão independentemente de nosso pensamento; ou nossa atividade mental tem papel constitutivo na realidade matemática e a existência dos números depende e é indissociável de nosso pensamento. A primeira resposta representa um compromisso com a tese realista (1), de que a mente está separada da realidade, pelo menos para o caso da realidade matemática, e a segunda resposta representa um compromisso com a tese oposta, que poderíamos chamar de *tese idealista*:

(2) *A mente não está separada da realidade.*

Posicionar-se com relação a este debate, no entanto, não nos compromete, necessariamente, com nenhuma tese ontológica sobre se os números existem ou não. Respondê-lo negando qualquer papel à mente na estrutura da realidade matemática deixa ainda sem resposta a questão sobre se os números existem ou não. Respondê-lo, ao contrário, atribuindo papel constitutivo à mente na estrutura da realidade matemática, aproxima-nos um pouco mais de um compromisso com a existência dos números. Mas não a tal ponto de tornar incoerente uma posição que assume a tese idealista (2) e ao mesmo tempo nega que o tipo de atividade mental relacionada com os números seja do tipo ontologicamente constitutivo.<sup>11</sup>

Como não se trata de um debate ontológico, mas metafísico, sobre o papel da mente na estrutura da realidade, o método do compromisso ontológico de Quine não nos ajudará. Assumir ou evitar compromissos ontológicos com os números não nos vinculará a nenhum dos lados deste debate. Será que há algo que nos vincula? Haveria algo, algum padrão ou indicador que faria no caso de debates metafísicos como este o mesmo papel que os compromissos ontológicos fazem no caso de debates ontológicos? Haveria algo que pudéssemos chamar de *compromisso metafísico* de um discurso ou teoria? Nossa aposta é a de que há. Da mesma forma que as afirmações existenciais das teorias expressam seus compromissos ontológicos, o sistema de lógica que tomamos como regulando uma determinada teoria ou discurso expressaria seus compromissos metafísicos.

A ideia aqui é que se aceitamos o fundamento da noção de compromisso ontológico, ou seja, a concepção de que “a existência é o que o quantificador existencial expressa” (Quine, 1969, 97), então devemos aceitar também que os princípios lógicos que regulam o comportamento dos quantificadores e demais conectivos são princípios metafísicos que exprimem um conceito de existência e delimitam os aspectos estruturais mais básicos da realidade. Este fato é o principal fundamento de nossa hipótese do isomorfismo e do

<sup>11</sup> O fundamento desta posição é o fato de que alguns de nossos pensamentos podem simplesmente não satisfazer os critérios que consideramos como suficientes para que eles sejam constitutivos do que tomamos como pertencente à realidade. Ainda que a mente crie realidade, isso não significa, necessariamente, que todos os nossos pensamentos casuais devam criar realidade. Pode ser que apenas os pensamentos que satisfaçam determinados critérios o façam. Com relação às entidades matemáticas esta seria uma posição bem estranha, já que não pensamos casualmente sobre elas, mas de um modo extremamente estruturado e regado pela matemática. Poderíamos nos perguntar: se nem nossos pensamentos matemáticos satisfazem os critérios de constituir a realidade, quais pensamentos satisfariam? No entanto, apesar de estranha, esta não é uma posição logicamente impossível.

critério da incompatibilidade lógica. Se as afirmações existenciais de uma teoria expressam as entidades com as quais esta teoria compromete-se ontologicamente, então a lógica que regula os aspectos mais gerais destas afirmações existenciais e todas as demais afirmações da teoria, a lógica que estabelece em que condições um conjunto de sentenças tem como consequência determinada afirmação (existencial ou universal ou de qualquer outro tipo) expressa os compromissos metafísicos desta teoria.

O próprio Quine não ignorava este fato. Em *Existence and quantification*, refletindo sobre as diferenças entre a lógica clássica e a intuicionista, ele, um tanto a contragosto, admite:

A teoria da quantificação clássica goza de uma extraordinária combinação de profundidade e simplicidade, beleza e utilidade. [...] Desvios dela, em contraste, tendem a parecer especialmente arbitrários. Mas como eles existem, o mais claro e simples é dizer que *conceitos de existência desviantes existem junto com eles*. (Quine, 1969, 112-113)

Neste mesmo texto, algumas páginas antes, Quine já havia afirmado que

[o] intuicionista tem uma *doutrina do ser* diferente da minha, na medida em que tem uma *teoria da quantificação* diferente; e eu simplesmente estou em desacordo com o intuicionista tanto num caso quanto no outro. (Quine, 1969, 108)

Quine admite explicitamente aqui que a diferença entre as teorias da quantificação clássica e intuicionista, ou seja, entre as lógicas de primeira ordem clássica e intuicionista constitui-se em uma diferença entre duas concepções distintas de existência, uma clássica e uma outra intuicionista. Assim a questão que talvez seja a mais central da metafísica, aquela que dá os traços mais gerais da estrutura da realidade ao abordar o significado e características da existência é simétrica à questão mais fundamental da lógica, sobre quais são os princípios lógicos universalmente válidos. Ao divergirem sobre quais são os princípios lógicos válidos, clássicos e intuicionistas divergem também, e por causa disso, nos modos como concebem a estrutura da realidade e entendem o próprio significado da existência.

Apesar deste reconhecimento, Quine simplesmente não estava disposto a debater metafísica nestes termos. Ele não estava interessado em testar as hipóteses metafísicas produzidas arbitrariamente e *a priori* pelos filósofos. Seu interesse no assunto limitava-se ao seu método ontológico naturalista descrito na seção anterior. A pauta da ontologia seria dada por nossas melhores teorias científicas que deveriam ser arregimentadas e avaliadas segundo a lógica clássica de primeira ordem. Então a doutrina metafísica de Quine, a única estrutura da realidade que ele se dispõe a aceitar é a que emana do (ou supervem ao) conceito de existência vinculado à lógica clássica de primeira ordem, que ele considerava adequado e suficiente para lidar com todas as questões ontológicas postas pelas teorias científicas. O método de Quine está limitado, então, ao que temos chamado de *interpretação ontológica* das questões metafísicas e a sua doutrina metafísica propriamente dita está limitada aos compromissos vinculados à lógica clássica de primeira ordem.

Mas qual exatamente é a doutrina metafísica vinculada ao conceito de existência dado pela lógica clássica de primeira ordem? Ao mesmo tempo que faz sentido afirmar que ao regular o comportamento mais geral dos quantificadores os princípios lógicos constituem-se em princípios metafísicos, faz igual sentido perguntar qual é a imagem metafísica que estes princípios nos dão. Afinal, os metafísicos não costumam na maioria das vezes apresentar suas propostas através de princípios lógicos de inferência, mas através de proposições mais imagéticas como as teses realista (1) e idealista (2), por exemplo.

Aqui a analogia com a geometria analítica proposta na Seção 2 novamente nos ajuda. Os princípios e teoremas lógicos seriam como as equações algébricas e a posição metafísica a eles vinculada seria como as figuras geométricas descritas pelas equações algébricas. À equação “ $x^2 + y^2 = 4$ ”, por exemplo, vincula-se uma circunferência, de centro  $(0, 0)$  e raio 2. Podemos, analogamente, perguntar qual é a posição metafísica que se vincula aos teoremas e princípios da lógica clássica de primeira ordem. De um modo mais geral podemos nos perguntar como extrair as diferentes teses metafísicas dadas pelos conceitos de existência de diferentes sistemas lógicos.

Infelizmente não temos uma fórmula geral que responde a todas as questões deste tipo. A analogia com a geometria analítica chega aqui ao seu limite. Equações algébricas vinculam-se individualmente às figuras geométricas dadas por seus gráficos cartesianos. Mas não temos a pretensão de vincular um a um os princípios de inferência com suas supostas imagens metafísicas. Não creio que haja uma tal vinculação um-a-um. Da mesma forma que Quine defende que os portadores de compromissos ontológicos não são as sentenças tomadas individualmente, mas as teorias (conjuntos de sentenças fechados pela relação de consequência lógica), defendendo igualmente que os portadores de compromissos metafísicos não são os princípios ou teoremas lógicos tomados individualmente, mas os sistemas lógicos. Lógicas distintas representariam conceitos de existência distintos e vinculariam-se a propostas metafísicas distintas sobre a estrutura da realidade.

A abordagem metodológica que sugiro, então, para esta tarefa de procurar quais são as posições metafísicas vinculadas aos diversos sistemas lógicos, consiste na análise lógica das propostas e debates metafísicos. O critério da incompatibilidade lógica propõe que uma divergência, um debate específico, será metafísico apenas quando as posições adversárias forem regidas por lógicas distintas. Devemos então vasculhar a literatura procurando pelos debates supostamente metafísicos e devemos fazer uma análise lógica das posições adversárias buscando identificar quais são os sistemas lógicos compatíveis com as inferências efetivamente empregadas em cada um dos lados de cada debate. Ao eventualmente encontrarmos um sistema lógico que seja compatível com a as inferências lógicas efetivamente utilizadas pelos defensores de uma proposta metafísica específica, poderemos vincular esta proposta metafísica explicitamente defendida ao sistema lógico compatível que encontramos. Ao fazer isso estaremos vinculando uma tese metafísica específica, defendida por um dos lados do debate, com um sistema lógico específico, o sistema compatível com as inferências lógicas efetivamente utilizadas pelos defensores desta tese metafísica. Ao aplicarmos esta abordagem a inúmeros debates poderemos, pouco a pouco, compor a rede de posições metafísicas vinculadas aos diversos sistemas lógicos.

Esta abordagem metodológica, no entanto, precisa ser utilizada com bastante cautela. Na próxima Seção discutiremos um pouco mais detidamente as razões desta cautela. Utilizaremos a proposta de Michael Dummett para o tratamento das diversas instâncias do debate entre realismo x idealismo tanto como um exemplo de aplicação desta abordagem metodológica, quanto como um caso de aplicação bem sucedida do critério da incompatibilidade lógica.

## 5. Realidade Diferente Demanda Lógica Diferente

Conforme sugerimos na Seção anterior, ao limitar-se à interpretação ontológica das questões metafísicas, o método de Quine compromete-se com a posição metafísica dada pela tese realista (1) de que a mente está afastada da realidade. Mas além disso, vimos também



que ao restringir-se à lógica clássica de primeira ordem, o método de Quine compromete-se também com a posição metafísica vinculada ao conceito de existência expressado pela lógica clássica de primeira ordem. Juntando então estes dois compromissos poderíamos concluir que uma das características da posição metafísica vinculada à lógica clássica de primeira ordem é a tese realista (1).

Se este for mesmo o caso, em qualquer debate metafísico, qualquer um que restrinja a lógica que rege sua posição à lógica clássica de primeira ordem estará, em conjunto, assumindo a tese realista (1) que afirma que a mente está separada da realidade. Mais importante ainda, qualquer um que discorde da tese realista (1) e queira atacá-la, terá que abdicar da lógica clássica de primeira ordem e necessitará adotar uma lógica alternativa. Michael Dummett, que por motivos completamente diversos aos de Quine também defendia que lógicas diferentes se configuram em posições metafísicas diferentes (Durante, 2011), apontou que os matemáticos construtivistas da escola de Brouwer foram os primeiros a perceber que para rejeitar uma posição metafísica realista eles precisariam, também, rejeitar a lógica clássica (Dummett, 1991, 9).

O nosso critério da incompatibilidade lógica para os debates metafísicos fundamenta-se tanto no critério de compromisso ontológico de Quine, e em sua vinculação da existência com a quantificação, quanto nas propostas de Michael Dummett para uma teoria do significado, que nos daria “as bases lógicas da metafísica”, título de seu livro seminal. Uma das principais preocupações metafísicas Dummett foi procurar uma metodologia unificada para o tratamento das diversas instâncias do debate entre as várias formas possíveis das teses realista (1) e idealista (2).<sup>12</sup>

Dummett considerava equivocada a interpretação ontológica destes debates. Por exemplo, segundo ele um debate entre um fenomenalista e um realista com relação ao mundo físico não é um debate ontológico restrito sobre se os objetos materiais existem ou não, mas sim um debate metafísico amplo sobre qual é o papel da experiência sensorial na constituição da realidade física. Nosso conhecimento sobre o mundo físico é dado pelos sentidos, por nossas experiências sensíveis. Para os realistas estas experiências sensoriais não teriam qualquer papel na estrutura da realidade. Elas estariam completamente separadas da realidade sobre a qual nos informam e seriam veículos neutros de informação sobre o mundo físico. A versão da tese realista neste caso seria:

(1') *A realidade está separada das experiências sensoriais.*

Já para os fenomenalistas as nossas experiências sensoriais não seriam veículos neutros de informação sobre o mundo físico, mas teriam papel constitutivo na realidade física, fazendo parte de sua estrutura. O fenomenalismo é uma versão do idealismo, e a tese idealista específica que ele suporta seria:

<sup>12</sup> Dummett preferia chamar de *antirrealismo*, ou de antirrealismos, as diversas teses que genericamente estou rotulando aqui de idealistas, das quais a tese (2) é apenas uma instância. Ele preferiu esta nomenclatura porque o termo idealismo é bastante carregado de interpretações vindas da história da filosofia com as quais ele queria evitar aproximações desnecessárias. Algumas das abordagens que ele classifica como antirrealistas, tais como o formalismo na filosofia da matemática, são, de fato, bastante distantes do significado tradicional de idealismo. Como ele considerava equivocada a abordagem tradicional às questões metafísicas e fez uma proposta metodológica reformista, ele preferiu utilizar uma nova terminologia. Optei, em contraste, por manter a terminologia tradicional justamente para ajudar a marcar o fato de que não tenho qualquer pretensão de que o critério da incompatibilidade lógica seja uma proposta metodológica reformista. Se bem-sucedido, o máximo que ele pode fornecer é um instrumento analítico para ajudar a interpretação e compreensão dos debates metafísicos, tanto os debates tradicionais quanto os contemporâneos. No contexto dos exemplos que utilizo aqui, os rótulos antirrealismo e idealismo podem, de um modo geral, ser entendidos como intercambiáveis sem maiores problemas.

(2') *A realidade não está separada das experiências sensoriais.*

Da mesma forma, a divergência entre platonistas e construtivistas sobre as entidades matemáticas não é uma questão ontológica sobre se os números (e demais entidades matemáticas) existem ou não, mas uma divergência metafísica sobre o papel da mente na constituição da realidade matemática.

Dummett considerava equivocada a interpretação ontológica porque ele não via como estes debates, tratados como questões ontológicas, poderiam ser solucionados. Ele não partilhava da confiança de Quine no método naturalista fundado nos compromissos ontológicos das teorias científicas e nem tampouco aceitava as propostas ontológicas tradicionais e *a priori* dos filósofos, que ele chamava de abordagens de cima para baixo (*top-down*). Ele defendia, em contraste, uma abordagem de baixo para cima (*bottom-up*) que seria obtida através da proposição de uma teoria do significado para a parte da linguagem ligada aos termos do debate. Esta teoria do significado seria feita sem quaisquer pressuposições metafísicas iniciais, e seria baseada apenas no papel inferencial das expressões linguísticas que nos seria dado através de uma análise de seus usos. Ao entender, através de uma análise do uso da linguagem, o papel inferencial das expressões relacionadas a um dado posicionamento metafísico, entenderíamos o significado dos termos envolvidos, através da lógica (teoria do significado) que rege estes discursos. Deste entendimento emergiria de um modo apenas superveniente uma imagem metafísica da situação. A metafísica, para Dummett, se limitaria a esta imagem superveniente que emerge da lógica e da teoria do significado.<sup>13</sup>

Os debates deveriam ser tratados caso a caso e uma posição realista (ou idealista) com relação a determinada instância do debate não comprometeria ninguém com a mesma posição realista (ou idealista) nas demais instâncias. Apesar de defender a independência das diversas instâncias do debate, em sua busca por uma unificação metodológica Dummett identificou que o que liga todas as abordagens realistas por um lado e todas as abordagens contrárias ao realismo por outro é uma divergência no conceito de verdade.

Seja qual for a concepção de verdade e de realidade que adotemos, aquilo que tomamos como verdadeiro deve ser uma expressão da realidade (Chateaubriand, 2001, 16). Se além disso, de acordo com a tese realista, a realidade está separada da mente (ou da experiência sensorial), então aquilo que é verdadeiro também está. Ou seja, a verdade não pode depender de nossa mente nem de nossas experiências sensoriais. Assim, para um realista, uma afirmação será verdadeira ou falsa independentemente de nossas capacidades mentais ou sensoriais de verificar sua verdade ou falsidade. Dummett mostrou que a principal característica que une todas as abordagens realistas é que de acordo com elas *a verdade é transcendente à verificabilidade*.

Para os realistas, então, uma determinada afirmação declarativa P deve ser verdadeira ou falsa, independentemente da possibilidade de haver ou não qualquer tipo de evidência contra ou a favor de P. O realismo, segundo esta concepção de Dummett, pode então ser entendido como uma defesa da neutralidade epistemológica da verdade. Para um realista a verdade ou falsidade de uma afirmação é independente de qualquer tipo de restrição que possa haver para o reconhecimento desta verdade ou falsidade.

Por outro lado, para um idealista (ou antirrealista na nomenclatura de Dummett), a mente ou a experiência sensorial não está separada da realidade, mas, ao contrário, faz parte de sua constituição e é indissociável dela. Então, sendo a verdade expressão da

<sup>13</sup> Uma boa apresentação desta proposta metodológica encontra-se na introdução de Dummett (1991), cujos demais capítulos a desenvolvem detalhadamente.

realidade, aquilo que é verdadeiro também estará indissociavelmente ligado à nossa mente ou às nossas experiências sensoriais e, por isso, não pode haver verdade ou falsidade onde elas não podem ser mentalmente ou sensorialmente reconhecidas. Ou seja, para um antirrealista não há verdade ou falsidade onde não há condições de verificabilidade. Segundo Dummett, então, a principal característica que une todas as abordagens antirrealistas é que de acordo com elas *a verdade depende da verificabilidade*.

Uma sentença  $P$ , então, só pode ser verdadeira, para um antirrealista, mediante evidência favorável a  $P$ , e só pode ser falsa mediante evidência contrária a  $P$ . Então o antirrealismo representa, para Dummett, uma defesa da dependência epistemológica da verdade. Para um antirrealista a verdade e o reconhecimento da verdade não são noções distintas. São, ao contrário, indissociáveis. Não haverá verdade ou falsidade onde não pode haver reconhecimento da verdade ou falsidade.

Vejam um rápido exemplo. Nós não sabemos com certeza se a afirmação “o universo é infinito” é verdadeira ou falsa. Há teorias divergentes sobre esta questão e não há evidências definitivas para assegurar nem sua verdade nem sua falsidade. Mais ainda, não sabemos se o tipo de evidência necessário para decidir a verdade ou falsidade desta afirmação poderá algum dia ser obtido. Ou seja, não há clareza sobre se é possível haver evidência para tal afirmação. Tomar estas incertezas sobre a verificabilidade da afirmação de que o universo é infinito como motivo suficiente para considerá-la nem verdadeira nem falsa é, segundo Dummett, adotar uma postura antirrealista com relação a questão da infinitude do universo. Por outro lado, admitir que a afirmação tem que ser verdadeira ou falsa independentemente da própria possibilidade de algum dia haver evidência favorável ou contrária a ela é, também segundo Dummett, adotar uma postura realista sobre esta questão.

Esta caracterização semântica de Dummett é compatível com nossa caracterização mais tradicionalmente metafísica ou imagética dos debates, representada pelas teses realista (1) e (1') e idealista (2) e (2'). Se aceitamos a tese realista e consideramos que a realidade do universo está separada de nossa mente ou sensibilidade, então ou ele é finito ou ele é infinito, independentemente de se há ou não possibilidade de algum dia obtermos evidência para o reconhecimento da verdade ou falsidade da afirmação de sua infinitude. Por outro lado, se aceitamos a tese idealista e consideramos que a realidade do universo depende de nossa mente ou sensibilidade, e se juntamente reconhecemos que a situação atual de nossas considerações científicas, filosóficas e empíricas sobre o universo não nos dá evidências suficientes para decidirmos sobre sua finitude ou infinitude, nem nos garante se algum dia poderá haver tais evidências, então devemos reconhecer também que a afirmação de que o universo é infinito não é nem verdadeira nem falsa, porque este ponto da realidade ainda não foi constituído por nossa sensibilidade ou pensamento, permanecendo um ponto aberto. Segundo a tese idealista não haveria, neste caso, qualquer fato na realidade sobre a infinitude ou finitude do universo e, portanto, também não haveria verdade ou falsidade vinculada à afirmação de que o universo é infinito.

Mas se para os realistas qualquer sentença declarativa  $P$  deve ser verdadeira ou falsa, independentemente da evidência, então para qualquer  $P$  a declaração

$$(P \vee \neg P)$$

que é o conhecido princípio lógico do terceiro excluído, será verdadeira. Afinal, quando  $P$  é verdadeira,  $\neg P$  é falsa e quando  $P$  é falsa,  $\neg P$  é verdadeira, e a disjunção de verdadeiro com falso é sempre verdadeira.

Por outro lado, se a verdade depende da verificabilidade, como os antirrealistas acreditam, então pode haver uma sentença  $Q$  que careça de evidência tanto favorável quanto contrária. Neste caso,  $Q$  não é nem verdadeira nem falsa. Mas qualquer que seja o motivo para a falta de evidência da verdade ou falsidade de  $Q$ , ele será também um motivo para não haver evidência para a verdade ou falsidade de  $\neg Q$ , que também não será nem verdadeira nem falsa. Como tanto  $Q$  quanto  $\neg Q$  não são nem verdadeiras nem falsas, sua disjunção  $(Q \vee \neg Q)$  também não será. Em particular,  $(Q \vee \neg Q)$  não será verdadeira. Ou seja, nem todas as instâncias do terceiro excluído serão verdadeiras e, por isso, o terceiro excluído não será um princípio lógico para o antirrealista (Dummett, 1978) (Durante, 2011).<sup>14</sup>

Então apenas para os realistas o terceiro excluído será um princípio lógico, uma declaração impossível de ser falsa, verdadeira em todas as circunstâncias. Os antirrealistas não têm motivos para aceitar a validade irrestrita do terceiro excluído. Na verdade, eles têm motivos para rejeitá-la. A concepção mais geral da estrutura da realidade que eles assumem, na qual a mente ou a sensibilidade têm parte constitutiva e a verdade depende da verificabilidade, lhes dá motivos para negar a validade universal do terceiro excluído. Mas se realistas e antirrealistas discordam sobre a validade de um princípio lógico, então os sistemas lógicos que regulam as inferências aceitas por cada grupo são também diferentes. Ao rejeitarem o terceiro excluído, os antirrealistas rejeitam também todos os argumentos cuja validade depende do terceiro excluído e os dois grupos terão, por isso, padrões de inferência incompatíveis dados por lógicas diferentes e incompatíveis (Durante, 2011, 39).

É importante neste ponto deixar clara a distinção entre a abordagem de Dummett e a nossa proposta do critério da incompatibilidade lógica. Dummett queria reformar a metafísica. Para ele qualquer proposta metafísica descrita em linguagem tradicional, tal como as teses (1) e (2), seria apenas uma imagem superveniente de uma teoria do significado cujas bases lógicas constituiriam a única e verdadeira proposta metafísica. Ele não estava interessado em entender e elucidar os debates metafísicos tradicionais, mas em substituí-los por uma análise lógica e semântica. Neste sentido, para entender qualquer instância do debate realismo x antirrealismo, ele precisaria produzir uma teoria do significado completa para a parte da linguagem que contém os termos do debate. Esta mostrou-se uma exigência bastante forte, e um fardo pesadíssimo. Acredito que o motivo desta grande dificuldade em utilizar de modo completo o método de Dummett está no fato de que as nossas pressuposições metafísicas são quase sempre incompletas e pouco detalhadas, deixando abertos muitos pontos que seriam necessários para a produção de uma completa teoria do significado nos moldes de Dummett. Ao não serem totalmente definidas e possuírem aberturas, as nossas ideias sobre metafísica que influenciam nosso uso da linguagem ficam sujeitas a serem completadas de modo diferente por filósofos diferentes, o que geraria desacordos nas possíveis teorias do significado. Talvez esta dificuldade

<sup>14</sup> Tanto no argumento realista para a aceitação do terceiro excluído, quanto neste argumento idealista para a sua rejeição, estou, obviamente, fazendo inferências lógicas. Alguém poderia, então, acusar a mim e a Michael Dummett de petição de princípio, afinal fazemos uso de princípios lógicos para justificar e criticar um princípio lógico. De um ponto de vista mais geral, não há saída para este dilema. Estamos condenados a fazer inferências lógicas em nossos argumentos, portanto, qualquer justificação de um princípio lógico será em algum sentido circular. No entanto, em nossa defesa neste caso específico, posso alegar que o próprio princípio lógico do terceiro excluído, cuja validade está sendo defendida por realistas e atacada por idealistas, não foi utilizado em nenhum dos dois argumentos. Utilizamos apenas princípios lógicos que poderíamos classificar de mais básicos, relacionados ao comportamento da negação e da disjunção em conexão com a presença e a ausência dos valores de verdade ‘verdadeiro’ e ‘falso’.

prática do método de Dummett explique porque sua abordagem, apesar de fecunda e interessante, tenha obtido tão poucos adeptos entre os filósofos contemporâneos.

Nossa proposta do critério da incompatibilidade lógica, no entanto, é bem menos pretensiosa. Nos exemplos de debates que temos apresentado, não descrevemos com detalhes as posições de cada um dos lados. Apenas mencionamos de um modo genérico os aspectos cruciais que esclarecem as divergências. No debate entre fenomenalismo e realismo sobre o mundo físico, por exemplo, o aspecto crucial é que um fenomenalista é alguém que defende a tese (2') de que nossas experiências sensíveis fazem parte constitutiva da realidade física e um realista entende, ao contrário, que as experiências sensíveis estão separadas da realidade física, defendendo a tese (1'). O que interessa é apenas esta divergência crucial, sendo os detalhes das diversas maneiras possíveis de completar uma posição fenomenalista ou realista dispensáveis para o debate específico. Portanto, dada a incompletude e parcialidade destas descrições das posições de cada lado do debate, não temos a pretensão de definir qual é a lógica que representaria todos os compromissos metafísicos de fenomenalistas ou realistas com relação ao mundo físico. Eu não sei nem preciso saber quais são todos os compromissos metafísicos dos fenomenalistas e dos realistas para poder aplicar o critério da incompatibilidade lógica de modo bem-sucedido. A única coisa que eu preciso saber para reconhecer que este é um debate metafísico legítimo é que seja qual for a lógica que expressa todos os compromissos metafísicos de um dos lados do debate, esta lógica é diferente e incompatível com a lógica que expressa todos os compromissos metafísicos do outro lado. E isso nós sabemos. Vimos que o terceiro excluído é uma sentença logicamente válida para os realistas, mas apenas contingente para os fenomenalistas. Então, seja qual for a lógica que expressa todos os compromissos metafísicos dos realistas, ela é incompatível com a lógica que expressa todos os compromissos metafísicos dos fenomenalistas.

O único requisito necessário para aplicar de modo bem-sucedido o critério da incompatibilidade lógica a uma divergência específica é, então, que encontremos alguma característica lógica da posição de um dos lados que seja incompatível com a posição do outro lado. Para as diversas instâncias do debate entre realismo x antirrealismo, Dummett nos mostrou que o princípio do terceiro excluído é uma destas características. As posições realistas aceitam o princípio do terceiro excluído, são compatíveis com ele. As posições antirrealistas, por seu turno, são incompatíveis com o terceiro excluído.

Devemos, no entanto, resistir à tentação de afirmar apressadamente que a lógica do realismo é a lógica clássica de primeira ordem e a lógica do idealismo é a lógica intuicionista. O que podemos afirmar com segurança é apenas que as teses idealistas são incompatíveis com a lógica clássica de primeira ordem e qualquer um que defenda uma posição idealista não pode aceitar a lógica clássica de primeira ordem como a lógica subjacente à sua posição. A lógica intuicionista de Heiting, alguma lógica intermediária (entre a intuicionista e a clássica), até mesmo a lógica modal S4 poderiam ser compatíveis com instâncias do idealismo. Analogamente, também podemos afirmar com segurança que as teses realistas são incompatíveis com a lógica intuicionista e com todas as lógicas compatíveis com a rejeição do terceiro excluído. Qualquer afirmação mais contundente neste sentido seria precipitada e desnecessária para uma aplicação bem-sucedida do critério da incompatibilidade lógica.

O exemplo mais bem desenvolvido destes debates entre realismo x antirrealismo é a divergência entre os matemáticos platonistas e construtivistas sobre o papel da mente na constituição da realidade matemática, que abordamos na Seção 4. Neste exemplo específico temos de um modo mais contundente a vinculação da lógica clássica de primeira ordem como a lógica que expressa todos os compromissos metafísicos da posição realista assumida

pelos matemáticos platonistas e a vinculação da lógica intuicionista como representando os compromissos metafísicos dos construtivistas. Este estado mais bem desenvolvido da explicitação das lógicas que representam os compromissos metafísicos de cada lado deste debate específico deve-se às características especiais da matemática, que tornam bastante explícitas as inferências lógicas empregadas em seu desenvolvimento. Mas mesmo aqui há margem para múltiplas interpretações. Não há acordo sobre qual versão da lógica intuicionista é a mais adequada para exprimir os compromissos metafísicos dos construtivistas, se a lógica intuicionista de Heyting, ou a lógica intuicionista minimal, um pouco mais restritiva, ou mesmo a lógica positiva ou outras versões mais restritivas ainda. De todo modo, na grande maioria dos debates metafísicos, as posições de cada lado não são tão detalhadas e precisas, e o máximo que podemos almejar é a identificação de alguma incompatibilidade lógica entre as posições adversárias que nos daria uma garantia de que as posições representadas por cada lado são, de fato, metafisicamente divergentes.

Exemplificamos, nesta Seção, um caso de aplicação bem-sucedida do critério da incompatibilidade lógica, e analisamos com algum detalhe quais os requisitos para uma tal aplicação e que informações ela nos dá. Vimos, através dos argumentos semânticos de Dummett sobre a divergência no conceito de verdade, que as diversas instâncias do debate entre realismo e antirrealismo, quando não são interpretadas de um modo ontológico restrito, configuram-se em genuínos debates metafísicos, porque as posições adversárias são logicamente incompatíveis. O lado realista aceita o terceiro excluído como uma validade lógica e o lado antirrealista o rejeita, tratando-o como uma sentença contingente que pode ser falsa em algumas circunstâncias. Na próxima Seção vamos criticar a tese tradicional de que haveria uma separação fundamental entre lógica e metafísica. De acordo com esta tese haveria uma separação entre os aspectos formais e materiais da realidade; a lógica seria responsável apenas pelos primeiros ao passo que a metafísica cuidaria exclusivamente dos segundos.

## 6. O Formal Encontra o Material: princípios lógicos como princípios metafísicos

O paradoxo do barbeiro é bastante conhecido e pode ser assim expresso: imagine uma pequena vila na qual mora e trabalha apenas um barbeiro que faz a barba de todos e apenas os habitantes que têm barba e não fazem a própria barba. Esta singela situação é paradoxal porque podemos fazer a seguinte pergunta capciosa: este barbeiro faz a própria barba? Se supusermos que ele faz, concluiremos que ele não faz, porque ao fazer a própria barba ele não satisfaz a condição que descreve quem são os habitantes que ele barbeia. Dissemos que ele barbeia apenas os que não fazem a própria barba. Por outro lado, se supusermos que ele não faz a própria barba, concluiremos que ele faz, porque ao não fazer a própria barba ele satisfaz a condição que descreve quem são os habitantes que ele barbeia. Dissemos que ele barbeia todos os que não fazem a própria barba. Temos, portanto, um paradoxo, porque qualquer tentativa de atribuímos verdade ou falsidade à afirmação “o barbeiro faz a própria barba” falha, pois, conforme expusemos, a suposição de verdade desta afirmação implica na sua falsidade e a suposição de falsidade implica na sua verdade.

Nathan Salmon<sup>15</sup> apresentou uma solução bastante interessante para o paradoxo do barbeiro. Ele disse que não há paradoxo do barbeiro porque um teorema lógico, que ele

<sup>15</sup> Em uma conferência proferida no IX Congresso Internacional de Filosofia Analítica, ocorrido em Fortaleza, em maio de 2014.

chamou de *lei de Russell*, garante que não pode existir nenhum barbeiro que faz a barba de todos e apenas os que não se barbeiam.

A lei de Russell é um teorema da lógica clássica de primeira ordem, ou seja, é um princípio lógico, uma declaração verdadeira em qualquer interpretação. Ela afirma que não há indivíduo  $x$  nem relação  $R$  em que  $x$  relaciona-se através de  $R$  com todos e apenas os indivíduos  $y$  que não se relacionam consigo mesmos através de  $R$ . Em linguagem formal:

$$\neg \exists x \forall y (R(x, y) \leftrightarrow \neg R(y, y))$$

Considere agora a seguinte relação binária interpretada:

$$\text{FazABarba}(x, y): x \text{ faz a barba de } y.$$

Se a lei de Russell é verdadeira para uma relação  $R$  qualquer, ela é em particular verdadeira para a relação **FazABarba**. Então a seguinte sentença, instância obtida da lei de Russell substituindo-se a relação genérica  $R$  pela relação específica e interpretada **FazABarba**, também é um teorema da lógica clássica de primeira ordem:

$$\neg \exists x \forall y (\text{FazABarba}(x, y) \leftrightarrow \neg \text{FazABarba}(y, y))$$

Mas esta sentença afirma precisamente que não há indivíduo que faz a barba de todos e apenas os indivíduos que não fazem a própria barba. Ou seja, a declaração de que não existe um barbeiro que faz a barba de todos e apenas os indivíduos que não se barbeiam é uma verdade lógica; um teorema da lógica clássica de primeira ordem que não pode ser falso. Então, afirma Salmon, as leis da lógica proíbem a existência do barbeiro do paradoxo. É logicamente impossível que haja um barbeiro que faça a barba de todos e apenas os que não se barbeiam, porque a declaração de sua não existência é um teorema lógico, uma sentença verdadeira em todas as circunstâncias possíveis. Não havendo tal barbeiro, não há também qualquer paradoxo que, pelos poderes da lógica clássica de primeira ordem, está solucionado.

Esta engenhosa solução do paradoxo do barbeiro proposta por Salmon assume dois compromissos indissociáveis:

- (i) Ela assume que seja qual for a lógica que rege nossas inferências sobre cidades, barbeiros e demais fatos corriqueiros da vida, esta lógica é compatível com a lógica clássica de primeira ordem, no sentido de que os teoremas da lógica clássica de primeira ordem também serão teoremas desta lógica.
- (ii) Ela também assume que aquilo cuja existência é proibida pela lógica clássica de primeira ordem, tal como o nosso barbeiro paradoxal, de fato não existe. Ou seja, aquilo cuja negação de sua existência é um teorema da lógica clássica de primeira ordem não existe mesmo.

O compromisso (i) é um compromisso com a lógica clássica de primeira ordem e o compromisso (ii) é um compromisso com o conceito de existência vinculado à lógica clássica de primeira ordem. Ao justificar a não existência do barbeiro afirmando que ela se segue de um teorema lógico, Salmon está utilizando este teorema como um princípio metafísico, que explicita uma característica fundamental de todos os seres. A saber: não há, não pode haver indivíduo que se relacione com todos e apenas os indivíduos que não se relacionam consigo mesmos. Esta é uma característica de todos os indivíduos e de todas as relações. É, por isso, um princípio metafísico.<sup>16</sup>

<sup>16</sup> O paradoxo do barbeiro é apenas uma versão lúdico-didática do paradoxo de Russell, que levou os matemáticos a abandonarem o esquema da compreensão irrestrita,  $\exists x \forall y ((y \in x) \leftrightarrow \varphi(y))$ , juntamente com toda a teoria de conjuntos ingênua. Se instanciamos ' $\varphi(y)$ ' por ' $\neg(y \in y)$ ', obtemos  $\exists x \forall y ((y \in x) \leftrightarrow \neg(y \in y))$ , cuja negação é uma instância da lei de Russell com ' $\in$ ' no lugar de ' $R$ '. Então nem todas as propriedades significativas exprimíveis em linguagem de primeira ordem definem um

De modo análogo, podemos considerar que todos os teoremas lógicos universais, com a forma

$$\forall x \alpha(x)$$

representam *necessidades* ou *obrigações* impostas a todos os seres, e os teoremas com a forma de negações de sentenças existenciais

$$\neg \exists x \alpha(x)$$

representam, como a lei de Russell, *impossibilidades* ou *proibições* a eles impostas. O conjunto dos teoremas com estas formas de uma dada lógica **L** caracteriza de modo bastante explícito o comportamento mais geral dos seres segundo esta lógica e, ao fazer isso, configura-se em uma detalhada explicitação metafísica da estrutura da realidade e do conceito de existência vinculados à lógica **L**.

Por exemplo, o teorema clássico

$$\neg \exists x (P(x) \wedge \neg P(x))$$

estabelece uma característica de todos os seres admissíveis por quem raciocina segundo a lógica clássica, que se constitui na proibição ou impossibilidade de qualquer indivíduo satisfazer e não satisfazer uma propriedade **P** qualquer. Já o teorema clássico

$$\forall y (P(y) \vee \neg P(y))$$

estabelece uma característica de todos os seres admissíveis sob a razão clássica que se constitui na obrigação ou necessidade de qualquer indivíduo satisfazer ou não satisfazer uma propriedade **P** qualquer.

Ao descreverem o que é impossível (ou proibido) e o que é necessário (ou obrigatório) a todos os seres, sem exceção, estes e os demais teoremas quantificacionais da lógica clássica constituem-se em uma caracterização da estrutura mais geral imposta a todos os seres, o que faz da lógica clássica uma doutrina metafísica.

Apresentada nestes termos lógicos, no entanto, esta doutrina metafísica carece dos apelos imagéticos das proposições metafísicas tradicionais. Os teoremas quantificacionais de um sistema lógico são como as equações algébricas que descrevem analiticamente uma situação geométrica cuja imagem superveniente só nos é dada à intuição pelas figuras obtidas nos gráficos cartesianos destas equações. As proposições metafísicas tradicionais corresponderiam às figuras geométricas dadas por estas imagens dos gráficos cartesianos. Mas do mesmo modo que o isomorfismo cartesiano mostrou que as equações algébricas e as figuras geométricas são apenas dois modos diferentes de apresentar e entender os mesmos fatos, a nossa hipótese do isomorfismo assume que as doutrinas metafísicas vinculadas aos sistemas lógicos são apenas um modo diferente, analítico, de apresentar teses metafísicas que, tradicionalmente, são apresentadas e discutidas do modo mais geométrico e intuitivo preferido pelos metafísicos.

Por um lado, as nossas práticas inferenciais tais como a exemplificada na solução de Nathan Salmon ao paradoxo do barbeiro corroboram o peso metafísico que damos aos teoremas quantificacionais da lógica. Mas por outro lado, os filósofos não costumam reconhecer ou rotular de metafísica a estrutura da realidade que os teoremas quantificacionais da lógica nos obrigam a aceitar. Eles, em geral, a rotulam de estrutura *formal* e tendem

---

conjunto. A lei de Russell impede que a propriedade “não é um membro de si mesmo”, que além de significativa é exprimível por ‘ $\neg(y \in y)$ ’, exista enquanto uma individualidade dada pelo conjunto de sua extensão. Esta restrição significa que propriedades não podem ser indiscriminadamente tratadas como coisas individuais que existem. A lei de Russell nos mostra que este fato não é uma limitação da teoria de conjuntos, mas uma característica da estrutura realidade dada pela lógica clássica, ou seja, um compromisso metafísico de todos os que raciocinam de acordo com a lógica clássica de primeira ordem.



a separar o domínio formal da lógica do domínio *material* da metafísica que seria o único ligado ao ser.

Esta separação pode ser explicada por dois fatores. O primeiro é a ubiquidade que a tese da neutralidade metafísica da lógica atingiu na ortodoxia contemporânea. E o segundo fator, relacionado ao primeiro, é que ao vasculharmos a história da filosofia veremos que a concepção dominante sempre foi a de que há uma única lógica, que se aprimorou ao longo do tempo, mas manteve-se até bem recentemente a única lógica. Já as propostas metafísicas sempre foram múltiplas e motivo de controvérsias desde a antiguidade. Havendo apenas uma lógica e muitas propostas metafísicas, não faria muito sentido rotular de metafísicas as exigências que esta lógica única impõe à estrutura da realidade. A lógica foi, então, tradicionalmente, vista como anterior à metafísica, neutra com relação às diversas controvérsias desta disciplina. E as exigências que a lógica faz à estrutura da realidade foram rotuladas de formais, em contraste com as exigências da metafísica que seriam materiais ou substanciais. Porém, a proliferação de propostas de sistemas lógicos alternativos ocorrida a partir da segunda metade do século XX deu-nos a oportunidade teórica de refletir sobre a natureza das diferentes exigências à estrutura da realidade feitas por diferentes sistemas lógicos e desafiar a sua classificação como exigências meramente formais. É isso que estamos fazendo neste artigo.

Podemos, por exemplo, refletir sobre o fato de que uma das exigências que a lógica clássica de primeira ordem faz à estrutura da realidade e que não é feita pela lógica intuicionista é que, segundo a lógica clássica, tudo o que existe deve satisfazer a seguinte obrigação:

$$\forall y (P(y) \vee \neg P(y))$$

A estrutura da realidade vinculada à lógica clássica é mais rígida do que aquela vinculada à lógica intuicionista, porque a primeira faz mais exigências do que a segunda. Todos os teoremas intuicionistas são teoremas clássicos, mas não o oposto. A fórmula acima, uma variação em primeira ordem do terceiro excluído, é teorema clássico, mas não é teorema intuicionista. Muitas outras fórmulas são teoremas clássicos, mas não intuicionistas. Então o conceito de existência dado pela lógica clássica exige de cada coisa que existe que ela satisfaça ou não satisfaça uma propriedade qualquer  $P$ . Já o conceito de existência dado pela lógica intuicionista não faz esta exigência. A estrutura da realidade da lógica intuicionista admite seres para os quais há certas propriedades que eles nem a satisfazem nem não a satisfazem. Então, a realidade da lógica intuicionista tem mais possibilidades que a realidade da lógica clássica. Sua estrutura é menos rígida, porque ela tem menos regras que a realidade descrita pela lógica clássica. Há fatos admissíveis aos seres concebidos segundo a lógica intuicionista que são inadmissíveis aos seres quando estes são concebidos segundo a lógica clássica. Se lembrarmos de nossas discussões da Seção anterior que relacionam o terceiro excluído às teses realistas e a rejeição do terceiro excluído às teses idealistas, esta estrutura menos rígida ligada à lógica intuicionista faz todo o sentido. Se a mente ou nossa sensibilidade tem papel constitutivo na realidade, conforme defendem os idealistas, então a realidade é mais maleável do que se ela fosse completamente independente de nossas capacidades. Se a mente, por exemplo, tem função constitutiva na realidade, podemos imaginar os seres como estando sujeitos a transformações ou processos possibilitados por nossa atividade mental que seriam proibidos aos seres quando sua existência é concebida como independente de nossa mente. A nossa mente poderia produzir fatos reais que seriam impossíveis ou inconcebíveis quando desvinculamos a existência de nossa atividade mental, como fazem os realistas.

Estas considerações sugerem que não há qualquer motivo *a priori* para separarmos as características estruturais da realidade dadas pelos teoremas lógicos, tais como  $\forall y (P(y) \vee \neg P(y))$ , como sendo de uma natureza formal e diversa das características estruturais da realidade dadas por princípios que temos historicamente chamados de metafísicos, como as teses realista e idealista. Os que defendem esta separação costumam apelar para as noções de possibilidade e necessidade. Dizem que há possibilidades lógicas que não são metafisicamente possíveis e que há necessidades metafísicas que não são logicamente necessárias.

Por exemplo, alguém poderia argumentar que seria logicamente possível, mas não metafisicamente possível, existir um indivíduo com o poder mágico de transformar qualquer objeto material em qualquer outro instantaneamente e sem esforço. Dizer que a existência deste indivíduo é logicamente possível significa dizer que a admissão de sua existência não implica nenhuma contradição lógica. Ao passo que dizer que a existência deste indivíduo não é metafisicamente possível significa dizer que a suposição de sua existência, apesar de não implicar nenhuma contradição lógica, contraria algum princípio que tomamos como princípio metafísico.

Porém, esta separação entre a natureza do que é lógico e do que é metafísico é arbitrária. Ela parte da pressuposição de que a lógica é anterior e mais ampla do que a metafísica. Mas não encontramos qualquer justificativa mais sólida para esta pressuposição além das frágeis razões históricas anteriormente mencionadas, relacionadas com a concepção largamente assumida no passado de que haveria apenas uma lógica mas muitas propostas metafísicas. Há, ao contrário, um motivo bastante razoável para negarmos esta separação e apostarmos na hipótese do isomorfismo entre lógica e metafísica. Ele se constitui no fato de que a principal característica tanto dos princípios tradicionalmente tidos como metafísicos quanto dos princípios lógicos, normalmente caracterizados como formais, é que estes dois tipos de princípios são absolutamente gerais. O aspecto mais básico que caracterizaria tanto a lógica quanto a metafísica seria a generalidade absoluta dos princípios destas duas disciplinas.

Na próxima Seção vamos argumentar em defesa desta tese da generalidade absoluta da lógica e, ao fazer isso, apresentaremos algumas motivações lógicas para a hipótese (do isomorfismo) de que os princípios lógicos se constituem em princípios metafísicos.

### 7. A Lógica Não Separa: generalidade absoluta

A definição tradicional da metafísica como a disciplina que lida com a caracterização da estrutura mais geral da realidade evidencia a sua generalidade absoluta. A estrutura mais geral da realidade é aquela da qual nada escapa. Porém, o que significa dizer que os princípios da lógica têm generalidade absoluta? Quando afirmo que os princípios da lógica têm generalidade absoluta, quero dizer com isso que os teoremas quantificacionais da lógica não dividem nem separam a realidade. Eles não se referem a seres específicos ou a parte dos seres. Eles sempre se referem a todos os seres. Quero dizer que não é possível afirmar nada particular através de um teorema quantificacional, mas apenas fazer afirmações absolutamente gerais que, justamente por isso, caracterizam a estrutura mais geral da realidade, configurando-se em princípios metafísicos.<sup>17</sup>

<sup>17</sup> A expressão “generalidade absoluta” costuma ser empregada na literatura relacionada ao domínio dos quantificadores e tem, neste contexto, um sentido diverso ao que uso aqui. Os detratores da tese da generalidade absoluta do domínio de quantificação assumem a posição de que é incoerente admitir que possa haver um domínio de quantificação absolutamente geral, que inclui absolutamente tudo, sem nenhuma restrição. Os domínios de todas as sentenças quantificacionais seriam sempre relativos a alguma

Não estou aqui me referindo a uma característica da lógica clássica ou de algum sistema lógico específico, mas a uma característica que considero fundamental a qualquer sistema lógico. Tão fundamental que serviria como um dos critérios necessários para a caracterização do que é lógica. Não pode haver, por isso, um teorema com demonstração que assegure a generalidade absoluta de qualquer sistema lógico. Podemos até enunciar e demonstrar um teorema similar, mas ele sempre será específico para algum sistema ou conjunto de sistemas lógicos particulares. Trata-se de uma questão de delimitação disciplinar e muitos lógicos poderão discordar desta proposta. O melhor que podemos fazer é apresentar argumentos e motivações para a aceitação deste critério, esperando que não estejamos sendo neste quesito tão arbitrários quanto aqueles que acusamos de arbitrariamente classificar a lógica como anterior e neutra com relação à metafísica.

Para além dos argumentos já apresentados nas Seções anteriores que nos encaminham à proposição da generalidade absoluta da lógica, vou reforçar a motivação para a sua adoção argumentando que a lógica de primeira ordem clássica, tanto quanto a intuicionista, satisfazem a generalidade absoluta. Dissemos acima que todos os teoremas lógicos com as formas

$$\forall x \alpha(x) \quad \text{e} \quad \neg \exists x \alpha(x)$$

são princípios metafísicos que estabelecem obrigações e proibições que caracterizam o comportamento mais geral dos seres. Mas e quanto aos demais teoremas lógicos, tais como os teoremas proposicionais? Qual sua relação com a metafísica?

Os teoremas de um sistema lógico relacionam-se uns com os outros, de modo que dado um teorema proposicional qualquer, sempre é possível encontrar versões quantificacionais dele que também serão teoremas. Os teoremas quantificacionais clássicos

$$\forall y (Q(y) \vee \neg Q(y)) \quad \text{e} \quad \neg \exists x (P(x) \wedge \neg P(x))$$

que apresentamos anteriormente são versões quantificacionais dos teoremas proposicionais clássicos

$$(Q \vee \neg Q) \quad \text{e} \quad \neg(P \wedge \neg P).$$

Este relacionamento entre os teoremas dos sistemas lógicos faz com que divergências nos teoremas proposicionais tenham consequências nos teoremas quantificacionais. A diferença entre a lógica clássica e a intuicionista é proposicional. Mas esta diferença proposicional tem consequências nos teoremas quantificacionais. Exatamente porque o terceiro excluído  $(Q \vee \neg Q)$  não é teorema da lógica intuicionista, esta sua versão quantificacional  $\forall y (Q(y) \vee \neg Q(y))$  também não será.

É justamente porque as divergências proposicionais têm consequências quantificacionais, que não precisamos limitar o critério da incompatibilidade lógica aos teoremas

---

classe restritiva. Já os defensores da tese da generalidade absoluta assumem que não há qualquer problema na admissão de um domínio de quantificação absolutamente genérico, que inclui absolutamente tudo (Santos, 2013). Meu uso da expressão “generalidade absoluta” neste texto não tem relação com este debate. Quando afirmo que uma determinada sentença quantificacional  $T$  é absolutamente geral (ou genérica), quero dizer com isso que para qualquer domínio de quantificação  $D$ , aquilo que  $T$  afirma aplica-se a todos os indivíduos do domínio  $D$  e não a apenas alguns. Então, se  $T$  é uma sentença absolutamente genérica, dado qualquer domínio de quantificação  $D$ ,  $T$  não separa os elementos de  $D$  entre os que a satisfazem e os que não a satisfazem. Ela é absolutamente geral porque é satisfeita por todos. O exemplo típico de uma afirmação absolutamente genérica é uma afirmação universal  $\forall x \alpha(x)$ . Mas há afirmações absolutamente gerais com outras formas lógicas, por isso a necessidade de um nome específico. Quando afirmo que um sistema lógico satisfaz a generalidade absoluta, quero dizer algo muito próximo da afirmação de que todos os seus teoremas quantificacionais são afirmações absolutamente genéricas, com alguns adendos que serão esclarecidos ao longo desta Seção.

quantificacionais. Para que uma lógica forneça um conceito de existência, é preciso que ela tenha quantificadores e seja pelo menos de primeira ordem. No entanto, como divergências proposicionais têm conseqüências quantificacionais, qualquer divergência lógica, mesmo que no nível proposicional, conecta-se a divergências quantificacionais e por isso caracteriza uma divergência metafísica. É também por causa desta interconexão entre os teoremas lógicos que afirmamos anteriormente que são os sistemas lógicos entendidos em sua totalidade e não os princípios lógicos isolados que expressam compromissos metafísicos.

Uma maneira bastante direta de demonstrar que um determinado sistema lógico **L** satisfaz o critério da generalidade absoluta seria demonstrar que **L** não tem afirmações existenciais entre seus teoremas. Ou seja, a generalidade absoluta de **L** estaria assegurada se nenhum de seus teoremas tivesse a forma  $\exists x \alpha(x)$ . Isso porque as sentenças existenciais são exatamente aquelas que dividem a realidade, que afirmam particularidades, especificidades de alguns seres, mas não de todos. Quando afirmo, por exemplo, que fantasmas existem, através da sentença  $\exists x \text{Fantasma}(x)$ , estou dividindo a realidade entre os indivíduos que a satisfazem e os que não a satisfazem, estou afirmando uma particularidade, destacando um aspecto específico da realidade e não uma característica geral. Ao afirmarem particularidades e especificidades aplicáveis apenas a alguns indivíduos, as afirmações existenciais não descreveriam as características mais gerais de todos os seres, não seriam absolutamente gerais. Então, se defendo que uma característica fundamental de qualquer princípio lógico é a sua generalidade absoluta, nenhum princípio lógico deveria ter a forma  $\exists x \alpha(x)$  e, portanto, nenhum sistema lógico deveria ter afirmações existenciais entre os seus teoremas.<sup>18</sup> O problema é que esta é uma suposição falsa.

A afirmação existencial  $\exists x (P(x) \vee \neg P(x))$ , por exemplo, é um teorema da lógica clássica de primeira ordem. Em geral, se  $\alpha$  é um teorema proposicional, então  $\exists x \alpha(x)$  também é teorema e, por isso, a lógica clássica está repleta de afirmações existenciais entre seus teoremas. Há, no entanto, uma esperança para esta situação. Quando  $\alpha$  é um teorema proposicional, não só  $\exists x \alpha(x)$  mas também  $\forall x \alpha(x)$  é teorema quantificacional. Ou seja:

$$\vdash \alpha \quad \Rightarrow \quad \vdash \exists x \alpha(x) \quad \Rightarrow \quad \vdash \forall x \alpha(x)$$

Então, as afirmações existenciais  $\exists x \alpha(x)$  que são teoremas porque a subfórmula no escopo do quantificador existencial  $\alpha(x)$  tem a forma lógica de um teorema proposicional  $\alpha$ , não representam qualquer problema para a generalidade absoluta da lógica clássica, já que a particularidade que elas afirmam  $\exists x \alpha(x)$  é devida a uma generalidade, uma vez que  $\forall x \alpha(x)$  também é um teorema.

Mas nem todas as afirmações existenciais que são teoremas da lógica clássica o são porque compartilham parte de sua forma lógica com um teorema proposicional. Em um sistema com identidade e constantes individuais, por exemplo, para qualquer constante 'c' a afirmação existencial  $\exists x (x = c)$  é um teorema, mas a afirmação universal correspondente  $\forall x (x = c)$  não é.

Este fato, no entanto, não ameaça a tese da generalidade absoluta da lógica. Apesar da forma existencial, a única particularidade da sentença  $\exists x (x = c)$  é linguística, dada

<sup>18</sup> Vale notar que as afirmações existenciais são precisamente a forma lógica das sentenças que apontam os compromissos ontológicos de uma teoria. Curiosamente, a forma lógica das afirmações que exprimem os compromissos ontológicos de uma teoria é a mesma forma lógica que não poderia tomar parte em uma caracterização metafísica da realidade, o que evidencia a distinção entre metafísica e ontologia que temos assumido. A ontologia nos dá as coisas que existem, e que preenchem completamente a realidade, cujas características gerais são providas pela metafísica que, ao fazer isso, define o significado da existência.

pela constante ‘c’. As sentenças com a forma  $\exists x(x = c)$  afirmam apenas que uma das convenções linguísticas da lógica clássica é que todos os nomes (as constantes individuais) da linguagem denotam algum indivíduo. Se nossa linguagem tiver  $n$  constantes individuais  $c_1, \dots, c_n$ , então para cada  $(1 \leq i \leq n)$ , a sentença  $\exists x(x = c_i)$  será um teorema que afirma que a constante ‘ $c_i$ ’ denota um indivíduo. Estas sentenças, tomadas em conjunto, exprimem a convenção linguística de que as constantes individuais devem denotar. Elas, então, apenas regulam o uso das constantes individuais. A única informação que elas nos dão, para além desta convenção de nomes, é a de que a realidade não é vazia. Há seres. Este fato, no entanto, não é particular, mas geral. Ao afirmar que há seres, que a realidade não é vazia, estamos afirmando uma característica geral da realidade assumida pela lógica clássica e, portanto, respeitando a tese da generalidade absoluta.

Um modo simples de perceber que os teoremas que envolvem constantes individuais não ameaçam a tese da generalidade absoluta é mostrar que se substituirmos cada constante individual que ocorre neles por uma variável ‘y’ e tomarmos o fecho universal da fórmula resultante, obteremos um teorema. O teorema  $\exists x(x = c)$ , por exemplo, nos dá o teorema  $\forall y \exists x(x = y)$ . De um modo geral, tanto na lógica clássica quanto na intuicionista de primeira ordem, o seguinte resultado

$$\vdash \alpha(c) \Rightarrow \vdash \forall y \alpha(y)$$

vale e sua demonstração em sistemas de dedução natural é um mero corolário do teorema de normalização de Prawitz (2006). Isso significa que nenhum teorema com a forma  $\alpha(c)$  ameaça a tese da generalidade absoluta, porque o que eles afirmam sobre ‘c’ é verdadeiro para qualquer indivíduo ‘y’ do domínio de quantificação, já que  $\forall y \alpha(y)$  também é teorema.<sup>19</sup>

Vimos, até agora, que para os teoremas existenciais que têm a forma de teoremas proposicionais, vale a seguinte regra:

$$\vdash \alpha \Rightarrow \vdash \exists x \alpha(x) \Rightarrow \vdash \forall x \alpha(x)$$

Se demonstrássemos que, independentemente da forma lógica de  $\alpha$ , dado um teorema existencial qualquer  $\exists x \alpha(x)$ , se ao substituirmos o quantificador existencial por um universal, a fórmula resultante  $\forall x \alpha(x)$  continua sendo um teorema, então teríamos um argumento completo para a defesa da generalidade absoluta da lógica. Ou seja, se demonstrássemos que, independentemente da forma de  $\alpha$ ,

$$\vdash \exists x \alpha(x) \Rightarrow \vdash \forall x \alpha(x)$$

este resultado garantiria que qualquer particularidade que pudéssemos afirmar através de um teorema existencial  $\exists x \alpha(x)$  seria apenas uma afirmação parcial de uma generalidade garantida por outro teorema, representado por  $\forall x \alpha(x)$ , o que manteria assegurada a generalidade absoluta da lógica clássica.

Este resultado não é válido em geral para a lógica clássica, mas há uma prova bastante simples dele para a lógica intuicionista, que também obtém-se como um corolário do

<sup>19</sup> Para os nossos propósitos de argumentar em favor da generalidade absoluta das lógicas clássica e intuicionista, a eliminação das constantes individuais dada pelo resultado ‘ $\vdash \alpha(c) \Rightarrow \vdash \forall y \alpha(y)$ ’ é suficiente e será assumida nos demais argumentos desta Seção. É interessante notar também que a eliminação das constantes individuais exigida pela arregimentação na notação canônica de Quine, que as substitui por descrições, não se constitui em um problema apenas quando nos restringimos à lógica clássica. Por não comportar condicionais contrafactuais, a lógica clássica se protege das célebres críticas de Saul Kripke (2001) à abordagem descritivista dos nomes, que só se aplicam em contextos que admitem tais condicionais. Já na lógica intuicionista, esta eliminação forçada das constantes é problemática, porque a famosa tradução de Godel-McKinsey-Tarsky entre a lógica intuicionista e a lógica modal S4 demonstra a capacidade da primeira de abrigar condicionais subjuntivos e contextos contrafactuais.

teorema de normalização.<sup>20</sup> A lógica intuicionista é, por isso, absolutamente geral. Todos os seus teoremas que contêm constantes individuais são meras afirmações parciais de teoremas universais sem constantes, e todos os seus teoremas existenciais são convertíveis em teoremas universais. Então não há, entre os teoremas da lógica intuicionista, nenhuma afirmação particular. Ela respeita a generalidade absoluta.

Na lógica clássica, na grande maioria dos casos em que  $\exists x \alpha(x)$  é um teorema,  $\forall x \alpha(x)$  também será, e as situações que desafiam a generalidade absoluta são bastante raras.<sup>21</sup> Mas há exceções. A única que conheço é a seguinte sentença:

$$\exists x (P(x) \rightarrow \forall y P(y))$$

Esta sentença é um teorema da lógica clássica de primeira ordem, mas a sentença  $\forall x (P(x) \rightarrow \forall y P(y))$ , obtida pela substituição do quantificador existencial principal por um quantificador universal, não é teorema clássico. Vale notar que, como este resultado vale para a lógica intuicionista, esta exceção e qualquer outra que possa haver é uma afirmação existencial que não é teorema intuicionista. Ou seja,  $\exists x (P(x) \rightarrow \forall y P(y))$  é um teorema clássico, mas não é um teorema intuicionista. Vamos comparar a afirmação existencial, que é teorema clássico, com a afirmação universal, que não é teorema, para tentar entender a situação.

$\exists x (P(x) \rightarrow \forall y P(y))$ : há pelo menos um indivíduo que, se ele for P, então todos são Ps.

$\forall x (P(x) \rightarrow \forall y P(y))$ : dado um indivíduo qualquer, se ele for P, então todos são Ps.

Pode até parecer que a primeira sentença está fazendo uma afirmação particular sobre um indivíduo peculiar, que ela está separando a realidade e que não é absolutamente genérica. Mas quando tentamos entender por que ela é um teorema, por que não pode haver nenhuma interpretação em que ela é falsa, percebemos que apesar de sua forma existencial, ela não faz uma afirmação particular.

A segunda sentença não é um teorema clássico porque podemos conceber a circunstância em que alguns indivíduos são Ps e outros não são. Nesta circunstância, quando

<sup>20</sup> Em Veloso et al. (2012) podemos encontrar vários resultados relacionados e mais gerais do que este nosso, que apresento agora o esboço da prova: suponha que temos uma prova normal,  $\pi$ , de  $\exists x \alpha(x)$  em um sistema de dedução natural para a lógica intuicionista. Sendo normal, e tendo  $\exists x \alpha(x)$  como sua última fórmula, a última regra de  $\pi$  será ou uma regra de introdução do existencial ( $\exists I$ ) ou uma regra do absurdo intuicionista ( $\perp$ ). Como a lógica intuicionista é consistente, a última aplicação de regra em  $\pi$  não pode ser ( $\perp$ ), caso contrário, se excluíssemos de  $\pi$  esta última regra obteríamos uma prova do absurdo sem premissas, o que seria uma prova da inconsistência da lógica intuicionista. Então a última aplicação de regra em  $\pi$  é uma aplicação de ( $\exists I$ ). Podemos substituir esta última aplicação de regra em  $\pi$  por uma aplicação da regra ( $\forall I$ ), que tem exatamente a mesma forma que ( $\exists I$ ), obtendo uma prova  $\pi'$  cuja última sentença será  $\forall x \alpha(x)$ . Esta aplicação de ( $\forall I$ ) respeita as restrições da regra, porque a premissa da última regra de  $\pi$  não depende de nenhuma hipótese, já que  $\pi$  é a prova de um teorema (não tem hipóteses abertas) e a aplicação de ( $\exists I$ ) não descarta hipóteses. A prova  $\pi'$  é, então, uma prova de  $\forall x \alpha(x)$  correta e sem hipóteses abertas. Logo  $\forall x \alpha(x)$  é teorema da lógica intuicionista.

<sup>21</sup> Como o conjunto de teoremas clássicos é infinito, falar de “grande maioria” sem nenhuma qualificação é um abuso de linguagem. Podemos entender “grande maioria” neste contexto como afirmando que para qualquer limite finito  $n$  no número de símbolos das fórmulas (complexidade), o número de teoremas existenciais com complexidade menor que  $n$  cuja substituição do quantificador existencial pelo universal resulta em um teorema universal é muito maior que o número de teoremas existenciais cuja mesma substituição resulta em uma fórmula universal que não é teorema.

instanciamos a sentença universal com um indivíduo que é  $P$ , teremos o antecedente do condicional  $P(x)$  verdadeiro, mas o conseqüente  $\forall y P(y)$  falso, já que estamos analisando a circunstância em que alguns indivíduos não são  $P$ . Então o condicional  $P(x) \rightarrow \forall y P(y)$  é falso e a sentença universal  $\forall x (P(x) \rightarrow \forall y P(y))$  também é, já que ela tem uma instância falsa.

A primeira sentença é um teorema porque tanto na circunstância em que alguns indivíduos não são  $P$ s, quanto na circunstância em que todos são, a sentença é verdadeira. Quando alguns indivíduos não são  $P$ s, se instanciamos o existencial com um destes indivíduos que não é  $P$ , teremos antecedente  $P(x)$  e conseqüente  $\forall y P(y)$  falsos, e portanto o condicional  $P(x) \rightarrow \forall y P(y)$  será verdadeiro e o existencial  $\exists x (P(x) \rightarrow \forall y P(y))$  também, já que tem instância verdadeira. Quando todos os indivíduos são  $P$ s, o conseqüente do condicional  $\forall y P(y)$  é verdadeiro e qualquer instância do existencial nos dá um antecedente do condicional  $P(x)$  também verdadeiro e portanto o condicional  $P(x) \rightarrow \forall y P(y)$  é verdadeiro e a sentença existencial  $\exists x (P(x) \rightarrow \forall y P(y))$  também, já que tem instância verdadeira.

Então, apesar de sua forma existencial, a sentença só é válida porque em cada uma das circunstâncias possíveis ela tem instâncias verdadeiras. As duas circunstâncias possíveis analisadas, a de que alguns indivíduos não são  $P$ s e a de que todos os indivíduos são  $P$ s esgotam as possibilidades. Encaradas isoladamente, cada uma destas circunstâncias representa uma situação particular, uma possibilidade específica e particular na qual os fatos podem se dar. Mas o conjunto destas duas possibilidades, ou seja, a afirmação de sua disjunção, não tem nenhuma particularidade e representa uma característica absolutamente geral da realidade, correspondendo a uma instância do terceiro excluído ( $\forall x P(x) \vee \neg \forall x P(x)$ ).

Acontece que o reconhecimento da validade da afirmação existencial  $\exists x (P(x) \rightarrow \forall y P(y))$  que apresentamos acima não nos informa qual destas duas circunstâncias possíveis é a responsável pela verdade da sentença. Há, então, dois motivos bastante distintos que justificam a verdade da sentença, e descobrir que a sentença é verdadeira não nos informa qual destes motivos é o que está em vigor. A sentença pode ser verdadeira porque alguns indivíduos não são  $P$ s e, portanto, ela pode ser instanciada de modo que antecedente e conseqüente do condicional sejam falsos. Mas ela pode também ser verdadeira porque todos os indivíduos são  $P$ s e portanto qualquer instância do existencial nos dá antecedente e conseqüente verdadeiros.

Apesar da sentença ter a forma lógica de uma afirmação particular que separa a realidade, como há duas circunstâncias distintas que representam dois modos distintos dela ser verdadeira, que esgotam todas as possibilidades em que os fatos podem se dar, e como o reconhecimento de sua verdade não nos diz qual dos dois é o modo pelo qual ela é verdadeira, então o reconhecimento de sua verdade não nos dá nenhuma informação particular sobre o mundo. Esta informação particular sugerida pela forma existencial da sentença só nos seria comunicada se pudéssemos saber qual das duas circunstâncias possíveis assegura a sua verdade. Como o simples fato de reconhecer a sentença como verdadeira não nos dá esta informação, e como a disjunção destas duas possibilidades é uma situação absolutamente genérica, pois esgota todos os casos, então podemos afirmar com segurança que, mesmo sendo uma afirmação existencial, sua verdade não nos informa nenhuma particularidade sobre a realidade e por isso não desafia generalidade absoluta da lógica.

Como a afirmação existencial  $\exists x (P(x) \rightarrow \forall y P(y))$  que estamos analisando não é um teorema da lógica intuicionista, podemos ter certeza de que não é uma peculiaridade do

argumento que apresentamos o fato de que o reconhecimento de sua validade não nos informa sobre qual de duas circunstâncias particulares possíveis é a responsável por sua verdade. Qualquer justificativa semântica para a validade desta sentença necessariamente dependerá de alguma instância do princípio do terceiro excluído, caso contrário ela seria um teorema intuicionista. Além disso, como demonstramos que na lógica intuicionista sempre que  $\exists x \alpha(x)$  é teorema,  $\forall x \alpha(x)$  também será, então qualquer outro suposto contraexemplo à generalidade absoluta que viermos a encontrar será semelhante a este, no sentido de que o reconhecimento de sua validade dependerá do princípio do terceiro excluído e, portanto, não desafiará a generalidade absoluta da lógica pelos mesmos motivos que apresentamos para o caso atual.

Podemos finalmente concluir que, mesmo possuindo certos teoremas existenciais cujas contrapartidas universais não são teoremas, a lógica clássica, tanto quanto a lógica intuicionista, respeita o princípio da generalidade absoluta. Nenhum de seus teoremas faz qualquer afirmação particular. Eles não dividem nem separam a realidade. Não se referem a características específicas de indivíduos específicos, mas apenas a características gerais de todos os indivíduos e constituem-se, por isso, em princípios metafísicos que nos dão uma caracterização da estrutura mais geral da realidade.

Conforme já comentamos, não é possível dar um argumento único para a generalidade absoluta de todos os sistemas lógicos, simplesmente porque não há consenso entre os lógicos sobre os limites do seu domínio de conhecimento. Além de argumentos como estes que apresentei sobre a generalidade absoluta de sistemas lógicos específicos, o máximo que podemos almejar é um resultado que nos informaria, por exemplo, quais as características formais genéricas que os sistemas lógicos deveriam ter de modo a respeitar a generalidade absoluta. Esta interessante tarefa está, no entanto, muito além dos propósitos deste texto.

Com a discussão dos debates entre realismo x idealismo apresentada na Seção 5, vimos um exemplo de aplicação bem-sucedida do critério da incompatibilidade lógica. Ou seja, uma aplicação do critério em que a incompatibilidade lógica das posições adversárias nos mostrou que a divergência ali tratada é uma divergência metafísica genuína. Apenas a título de exemplo, antes de passarmos às considerações finais vamos, na próxima Seção, discutir brevemente um caso em que a aplicação do critério da incompatibilidade lógica demonstra que a posição dissidente defendida por um dos lados não é, nos termos em que foi proposta por seu autor, uma dissidência metafísica legítima, caracterizando a divergência como meramente verbal.

### 8. McDaniel x van Inwagen: a univocidade do ser

Em “Ways of Being”, Kris McDaniel (2009) apresenta uma defesa do que ele considera ser a versão heideggeriana da tese de que a existência não é unívoca. Haveria, para Heidegger, múltiplos modos de ser: “there are many things which we designate as ‘being’, and we do so in various senses” (Heidegger, 1993, 26). McDaniel aponta rapidamente algumas raízes históricas da tese e propõe uma interpretação para a versão heideggeriana na qual os múltiplos modos do ser são identificados através da equivocidade de significados presente em nossos usos de expressões como “ser”, “existência”, “existe”, “é uma entidade”, “há”. Ele classifica estas expressões como *analógicas*, e define: uma expressão é analógica quando “ela tem um sentido genérico que, a grosso modo, aplica-se a objetos de diferentes tipos em virtude destes objetos exemplificarem características bastante diferentes” (McDaniel, 2009, 295). Ele apresenta a expressão mereológica “é uma parte de” como um modelo paradigmático de expressão analógica e exemplifica alguns de seus usos:



esta mão é uma parte deste homem, a classe das mulheres é uma parte da classe dos seres humanos, esta sub-região é uma parte do espaço, este minuto é uma parte desta hora, esta premissa é uma parte deste argumento, e assim por diante. (McDaniel, 2009, 295)

Para McDaniel há uma ideia comum a estes vários usos, mas não é esta ideia comum que explica e justifica cada um deles. Cada uso específico aplica-se a objetos de tipos muito distintos e é explicado por uma relação fundamental específica. Cada uma destas relações fundamentais representa um conceito distinto do que é ser uma parte que é particular ao tipo de coisas às quais ela se aplica. Portanto, os motivos pelos quais há partes em cada caso são também fundamentalmente diversos (p. 296). Ou seja, “ser uma parte de” é uma expressão analógica para McDaniel porque haveria diversos modos muito variados em que algo é parte de algo. Da mesma forma, nossos múltiplos usos das expressões ligadas aos seres exprimiriam o fato de que há muitos modos de ser, e o conceito de ser também seria analógico. Usarei a expressão ‘*tese da equivocidade*’ para me referir à posição de McDaniel.

Na outra ponta da divergência que nos interessa aqui está Peter van Inwagen, que defende a tese contrária de que a existência é unívoca, não havendo múltiplos modos de ser, mas um único. Para van Inwagen, o conceito de ser não seria analógico e qualquer coisa que exista, existe exatamente do mesmo modo e no mesmo sentido do que qualquer outra. Usarei a expressão ‘*tese da univocidade*’ para me referir à posição de van Inwagen.

A divergência entre as teses da univocidade e da equivocidade é, aparentemente, uma divergência metafísica sobre se na estrutura geral da realidade há lugar para diferentes modos de existir, ou se a realidade comporta apenas um modo de existir. Nosso objetivo aqui é aplicar o critério da incompatibilidade lógica para mostrar que, do modo específico como estes dois autores a tratam, esta não é uma divergência metafísica legítima, mas apenas um desacordo meramente verbal entre os adversários. Mas neste caso específico, o critério nos ajudará também a perceber que a causa deste desacordo verbal é um mau uso das palavras feito pelos defensores da tese da equivocidade. Veremos que o vocabulário que McDaniel importa de Heidegger sobre os muitos modos de ser é inadequado e enganoso.

Feitos estes esclarecimentos preliminares, apresentarei, a seguir, um breve resumo da tese da univocidade conforme defendida por van Inwagen (2009) em “Being, Existence, and Ontological Commitment” e os principais pontos da tese da equivocidade, defendida por McDaniel (2009) em “Ways of Being”. Em seguida apresentarei a crítica à tese da equivocidade que o próprio van Inwagen (2014) faz, em “Modes of Being and Quantification”, onde ele realiza uma arregimentação formal da tese da equivocidade na lógica clássica de primeira ordem, a mesma lógica em que ele arregimenta a sua própria tese da univocidade. Isso torna possível empregar o critério da incompatibilidade lógica e concluir que o desacordo não é genuinamente metafísico. Ao final esclarecemos quais as conclusões que podemos tirar desta análise.

Em “Being, Existence, and Ontological Commitment”, van Inwagen apresenta o modo como ele entende que a ontologia deveria ser praticada e quais suas teses fundamentais. Ele afirma que estas teses representam o modo como Quine entende as questões ligadas ao ser e as rotula de “5 teses da meta-ontologia de Quine”:

- T1) O ser não é uma atividade.
- T2) Ser e existência são o mesmo.
- T3) A existência é unívoca.

- T4*) O único sentido do ser ou existência é adequadamente capturado pelo quantificador existencial da lógica formal.
- T5*) As questões ontológicas devem ser resolvidas através do método do compromisso ontológico.

Se ser e existência são o mesmo, então falar sobre univocidade ou equivocidade do ser, ou sobre um único ou múltiplos modos de existir é a mesma coisa. Além disso, se a existência é unívoca e o seu único sentido é capturado pelo quantificador existencial, então há um único e irrestrito domínio de quantificação que abrange tudo o que há, e as regras lógicas, que regulam o quantificador existencial, esgotam as possibilidades para o significado da existência. Estas teses o autorizam a concordar com a afirmação de Quine de que “a existência é o que o quantificador existencial expressa. Há coisas do tipo  $F$  se e somente se  $\exists x F(x)$ ” (Quine, 1969, 97).

Além disso, como, de acordo com *T5*, van Inwagen aceita o método ontológico de Quine que apresentamos na Seção 3, então a lógica formal a que ele se refere na tese *T4* é a lógica clássica de primeira ordem. Ou seja, o quantificador existencial que captura o sentido único da existência é para van Inwagen o existencial da lógica clássica de primeira ordem.

A tese da univocidade, de acordo com van Inwagen, resume-se, então, na aceitação de que é na lógica clássica de primeira ordem que devemos arregimentar nossas teses sobre o ser, que há um único e irrestrito domínio de quantificação, que é acessado pelo quantificador existencial, cujas regras capturam o sentido único da existência.

Voltemo-nos agora para a tese da equivocidade. Vou me concentrar apenas nos pontos cruciais suficientes para o entendimento do debate, omitindo vários de seus aspectos e detalhes que não são relevantes para a nossa questão. Após as explicações iniciais sobre o significado mais geral da ideia da equivocidade do ser e de vários esclarecimentos paralelos, McDaniel esforça-se por explicar a posição de Heidegger sobre a equivocidade através de recursos formais como predicados, quantificadores e variáveis. Ele afirma que apesar de Heidegger defender que “o ser vem em muitos sabores”, ele reconhece que há um sentido genérico de ser. E afirma que “o sentido genérico de ser é representado na lógica formal pelo quantificador irrestrito” padrão da lógica de primeira ordem (McDaniel, 2009, 301).

Então ele se pergunta: “qual é a melhor forma de representar os diversos sentidos restritos do ser segundo Heidegger?” Após rapidamente recusar o uso de constantes individuais para esta tarefa, ele considera a possibilidade de predicados especiais representarem os diversos modos de ser. Mas também aqui descarta rapidamente esta possibilidade: “isto parece inapropriado, uma vez que este procedimento trata a atribuição de um modo de ser a uma coisa como a predicação de uma propriedade desta coisa”. E continua: “os modos do ser não são meras propriedades especiais que algumas entidades têm e que outras carecem, e por isso não são representados do modo mais perspicaz por predicados” (p. 302). E faz, na sequência, sua sugestão crucial:

uma ideia natural, então, é que os sentidos específicos de ‘ser’ são melhor representados por quantificadores. A noção de um quantificador restrito — um que varia sobre apenas algum subconjunto próprio do que o quantificador irrestrito varia — é perfeitamente inteligível. Os sentidos do ‘ser’ de Heidegger são propriamente representados em um sistema formal por quantificadores restritos. [...] Então para cada tipo especial de ser reconhecido por Heidegger, corresponde um quantificador restrito cujo domínio é uma subclasse própria do domínio do quantificador irrestrito, e que varia sobre todas e apenas aquelas coisas que têm este tipo de ser (p. 302).

Logo em seguida ele faz a seguinte estranha declaração.

Então representar os modos de ser de Heidegger através de quantificadores restritos — quantificadores que *em virtude de seu significado* variam apenas sobre algum subconjunto próprio do que o quantificador existencial irrestrito varia — parece um modo excelente de proceder. (pp. 302-303)

A estranheza aqui está na afirmação de que os quantificadores restritos são restritos em virtude de seu significado. A pergunta imediata que esta afirmação suscita é: o que é o significado de um quantificador? Onde, por quem e como ele é estabelecido? A única indicação de alguma resposta para esta questão que seu texto traz é uma citação de Eli Hirsch (2005, 76), em que ele afirma que os quantificadores seriam semanticamente restritos quando “regras semânticas implícitas na linguagem” estabelecem a restrição de seu domínio (McDaniel, 2009, 303). Mas McDaniel ignora esta sugestão e deixa completamente sem resposta a questão sobre como os quantificadores adquirem o significado que eles têm.

Mas continuemos por ora. Ele então propõe, para exemplificar, dois tipos de quantificadores restritos, o que ele chama de ‘*existenzial*’, cujo domínio de quantificação são as entidades cujo modo de ser é ter *existenz*, que é o modo de ser das coisas concretas e particulares. O outro quantificador é o que ele chama de ‘*subsistentzial*’, cujo domínio são as entidades que têm *subsistência* como o seu modo de ser, que é o modo de ser das coisas abstratas. E ele insiste que é em virtude do significado de cada um destes quantificadores que seu domínio de quantificação é especificado (p. 303).

Temos então três tipos de quantificadores existenciais, o quantificador existencial genérico da lógica clássica, cujo domínio de quantificação é absolutamente genérico, e os quantificadores restritos *existenzial* e *subsistentzial*, com seus domínios de quantificação restritos aos tipos de seres concretos e abstratos a que eles respectivamente se ligam.

McDaniel alega que, de acordo com a perspectiva de Heidegger, os quantificadores restritos *existenzial* e *subsistentzial* devem ser prioritários em significado com relação ao quantificador existencial genérico (p. 303). Então ele afirma que, sendo prioritários em significado em relação ao quantificador irrestrito, os quantificadores restritos devem ser semanticamente primitivos, e passa a investigar como usar estes quantificadores para entender de que modo o sentido genérico do ser, que é subsidiário dos sentidos restritos, poderia deles ser obtido. Para esta discussão ele apela ao conceito de fundamentalidade.

Esta discussão, do modo que ele a apresenta, me parece apressada e mal formulada. Sua conclusão é também meramente negativa. Ele não apresenta nenhum argumento positivo para suportar sua posição sobre a prioridade de significado dos quantificadores restritos relativamente ao genérico, mas apenas uma defesa a um argumento contrário a esta tese. Ainda que esta defesa seja bem sucedida, o problema de sua abordagem que propiciou a crítica de van Inwagen é o fato de que, para além da mera afirmação de que há diferentes modos de ser, e da nomeação de alguns destes modos, ele não diz uma palavra sequer que explique, demonstre ou esclareça em que consiste a distinção dos múltiplos modos de ser. Se a existência abstrata dos seres que subsistem é diferente da existência concreta dos seres que existem, em que consiste esta diferença? Para além de pertencerem a domínios de quantificação distintos, quais as características que distinguem estes dois domínios de quantificação? O que é permitido (ou possível) a seres de um tipo que é proibido (ou impossível) a seres de outro? Qual a motivação para esta separação entre os diferentes modos de ser e em que ela consiste? Não há uma palavra sequer esclarecendo estas questões no texto de McDaniel.

Esta ausência de qualquer esclarecimento sobre de que modo os múltiplos modos de ser são diferentes uns dos outros propiciou combustível para a crítica da tese de McDaniel da equivocidade do ser que van Inwagen (2014) elaborou em “Modes of Being and Quantification”. Ele aproveitou-se desta indefinição e propôs variadas maneiras de arregimentar a tese da equivocidade na lógica clássica de primeira ordem. Algumas utilizando uma versão polissortida da lógica de primeira ordem e interpretando os modos de ser através de quantificadores restritos, exatamente de acordo com as prescrições de McDaniel (van Inwagen, 2014, 20). Algumas outras, no entanto, formalizando os modos de ser não como quantificadores, mas como predicados cujas extensões coincidiriam com os seres de cada suposto modo (van Inwagen, 2014, 9).<sup>22</sup>

O ponto central de sua argumentação foi mostrar que qualquer argumento que fosse informalmente válido (ou inválido) quando avaliado de modo a respeitar a tese da equivocidade, continuaria válido (ou inválido) em qualquer de suas arregimentações na lógica clássica de primeira ordem. Se a este fato acrescentarmos a sua abordagem à ontologia, na qual o sentido único da existência é capturado pelo quantificador existencial geral da lógica clássica, então estas suas arregimentações representam uma prova de que a tese da equivocidade não afirma nada substancial, já que ela é compatível com a tese da univocidade. Ao possuir arregimentações na lógica clássica de primeira ordem que se comportam exatamente como a versão original não arregimentada, a tese da equivocidade se mostra compatível com a tese da univocidade, uma vez que o quantificador existencial único da lógica clássica exprime e representa a própria tese da univocidade, de acordo com van Inwagen. O fundamento da crítica de van Inwagen é, então, o caráter inerte da proposta de McDaniel.

Van Inwagen não menciona explicitamente nada semelhante ao nosso critério da incompatibilidade lógica, nem tampouco compromete-se com nossa hipótese do isomorfismo entre lógica e metafísica. Mas podemos “sentir o cheiro forte” do critério da incompatibilidade lógica aqui. Ele entende sua própria argumentação como uma vitória da tese da univocidade sobre a equivocidade, vitória esta propiciada por suas arregimentações compatíveis com a versão original não arregimentada da equivocidade com relação aos argumentos que elas validam ou invalidam.

A nossa interpretação da situação, dada pelo critério da incompatibilidade lógica é, no entanto, suavemente diferente. Nós não precisamos, em princípio, assumir a tese inwageniana de que um dos compromissos metafísicos da lógica clássica de primeira ordem é a univocidade do ser. A controvérsia sobre a generalidade absoluta do domínio de quantificação da lógica clássica presente na literatura (Rayo and Uzquiano, 2009) nos sugere cautela neste quesito. Na ausência desta suposição, as arregimentações de van Inwagen demonstram apenas que as teses da univocidade e da equivocidade são logicamente compatíveis e, portanto, pelo critério da incompatibilidade lógica, a divergência entre elas não corresponde a uma divergência metafísica genuína sobre a estrutura mais geral da realidade. É esta a informação principal que a aplicação de nosso critério a este debate nos dá. No entanto, quando colocamos lado a lado o discurso e vocabulário heideggeriano ligado à tese da equivocidade e o discurso e vocabulário quineano da tese da unicidade, e os confrontamos com os compromissos metafísicos mais óbvios que a lógica clássica de primeira ordem parece assumir, fica muito difícil discordar de van Inwagen.

---

<sup>22</sup>Como a lógica clássica polissortida é traduzível na lógica clássica com um único tipo de variável, então todas as arregimentações de van Inwagen da tese da equivocidade de McDaniel são equivalentes e compatíveis com a lógica clássica de primeira ordem tradicional.

Talvez haja razões para rejeitar também a tese da univocidade de van Inwagen, por causa de argumentos ligados à generalidade absoluta dos quantificadores ou outro motivo qualquer. No entanto, mesmo que venhamos a descobrir que a existência não é unívoca, a presente análise nos autoriza a afirmar com segurança que ainda que haja múltiplos modos de ser, não será a versão de McDaniel da tese da equivocidade de Heidegger que nos fornecerá a explicação deste fato.

### 9. Considerações Finais

A proposição do critério da incompatibilidade lógica foi motivada por dois tipos de preocupações: metafísicas e lógicas. Do lado da metafísica, minha preocupação primeira foi defender a disciplina dos ataques deflacionistas que ela sempre sofreu tanto do senso comum, quanto de cientistas e também de filósofos. Mas do mesmo modo que considero injustificados estes ataques, também sempre me incomodou o fato de que muitas propostas e debates metafísicos específicos presentes na literatura parecem dar razão aos deflacionistas. Refletir sobre a estrutura geral da realidade é uma atividade arriscada, escorregadia, traiçoeira e em muitos casos específicos os filósofos parecem tropeçar em suas próprias palavras e perderem-se em considerações incoerentes, intuições esdrúxulas e debates estéreis.

Do lado da lógica, sempre me intrigou o fato que todas as nossas justificativas lógicas se fundam em princípios que não são, eles próprios, justificados. O que fundamentaria os princípios lógicos? Qual a razão para o fato de que quando os aplicamos corretamente eles nos levam a conclusões confiáveis?

A hipótese do isomorfismo entre lógica e metafísica e o critério da incompatibilidade lógica para as divergências metafísicas representam a minha resposta a estas inquietações. Os princípios lógicos são confiáveis porque eles configuram-se em uma expressão da estrutura geral da realidade. São prescrições para o pensamento correto, conforme nos ensina Frege, porque são leis que estabelecem o que é, em seu aspecto mais geral. *Os princípios lógicos, são lógicos, porque são princípios metafísicos.*

As propostas metafísicas, por seu turno, enquanto caracterizações da estrutura mais geral da realidade, são, por isso, imagens supervenientes dos princípios lógicos. Vinculam-se a eles tanto quanto as figuras geométricas vinculam-se aos gráficos cartesianos de suas equações algébricas. Então aquelas propostas supostamente metafísicas, mas que não possuem qualquer especificidade lógica, não são genuinamente metafísicas. E as propostas genuinamente metafísicas não são estéreis, são, ao contrário, férteis. Porque ao possuírem especificidade lógica elas se espalham e influenciam todos os demais assuntos sobre os quais raciocinamos. Ao influenciarem e afetarem todos os assuntos sobre os quais raciocinamos, elas assumem o papel central que se espera de uma caracterização geral da realidade. *As propostas metafísicas, são metafísicas, porque são lógicas.*

## Referências Bibliográficas

- Aristotle and Richard McKeon. *The Basic Works of Aristotle*. Random House, New York, 1941.
- Francesco Berto and Matteo Plebani. *Ontology and Metaontology: a contemporary guide*. Bloomsbury Publishing, London, 2015.
- Rudolf Carnap. *The Logical structure of the world: pseudoproblems in philosophy*. University of California Press, Berkeley and Los Angeles, 1969.
- Oswaldo Chateaubriand. *Logical Forms*, volume 34 of *Coleção CLE*. UNICAMP, Centro de Lógica, Epistemologia e História da Ciência, Campinas, 2001. Part I - Truth and Descriptions.
- Itala D'Ottaviano and Hércules Feitosa. Sobre a história da lógica, a lógica clássica e o surgimento das lógicas não-clássicas. 2003. URL <ftp://ftp.cle.unicamp.br/pub/arquivos/educacional/ArtGT.pdf>.
- Michael Dummett. Realism. In *Truth and Other Enigmas*, pages 145–165. Harvard University Press, Cambridge, 1978.
- Michael Dummett. *The Logical Basis of Metaphysics*. Harvard University Press, Cambridge, 1991.
- Daniel Durante. Logic is metaphysics. *Principia*, 15(1):31–42, 2011.
- Daniel Durante. Compromisso ontológico. In J. Branquinho and R. Santos, editors, *Compêndio em Linha de Problemas de Filosofia Analítica*, pages 1–39. LanCog - Universidade de Lisboa, 2014.
- Hartry Field. *Science Without Numbers*. Oxford University Press, Oxford, 2016.
- Gottlob Frege and Montgomery Furth. [*Grundgesetze der Arithmetik*.] *The basic laws of arithmetic; translated and edited, with an introduction, by Montgomery Furth*. Cambridge University Press, Berkeley and Los Angeles, 1967.
- Martin Heidegger. *Being and Time*. Harper Collins Press, New York, 1993.
- Eli Hirsch. Physical-object ontology, verbal disputes, and common sense. *Philosophy and Phenomenological Research*, 70(1):67–97, 2005.
- Edmund Husserl. *Logical investigations*. Routledge, Bury St Edmunds, 2015.
- Saul A. Kripke. *Naming and Necessity*. Harvard University Press, Cambridge, 2001.
- Kris McDaniel. Ways of being. In David Chalmers, David Manley, and Ryan Wasserman, editors, *Metametaphysics: new essays on the foundations of ontology*, pages 290–319. Oxford University Press, Oxford, 2009.
- Dag Prawitz. *Natural Deduction: a proof-theoretical study*. Dover Publications, Mineola, N.Y., 2006.
- Willard van O. Quine. *From a Logical Point of View*. Harper Torchbooks, New York, 1963a.
- Willard van O. Quine. Logic and The Reification of Universals. In *From a Logical Point of View* Quine (1963a), pages 102–129.
- Willard van O. Quine. On what there is. In *From a Logical Point of View* Quine (1963a), pages 1–19.

- 
- Willard van O. Quine. Existence and Quantification. In *Ontological Relativity and Other Essays*, pages 91–113. Columbia University Press, New York, 1969.
- Agustín Rayo and Gabriel Uzquiano. *Absolute generality*. Clarendon, Oxford, 2009.
- Gonçalo Santos. Numbers and Everything. *Philosophia Mathematica*, 21(3):297–308, 2013.
- Robin Smith. Aristotle’s logic. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, summer 2019 edition, 2019.
- Tuomas E. Tahko. *An Introduction to Metametaphysics*. Cambridge University Press, Cambridge, MA, 2015.
- Alfred Tarski. The semantic conception of truth and the foundations of semantics. *Philosophy and phenomenological research.*, 4(3), 1944.
- Peter van Inwagen. Being, Existence, and Ontological Commitment. In David Chalmers, David Manley, and Ryan Wasserman, editors, *Metametaphysics: new essays on the foundations of ontology*, pages 472–506. Oxford University Press, Oxford, 2009.
- Peter van Inwagen. Modes of being and quantification. *Disputatio*, 6(38):1–24, 2014.
- Paulo Veloso, Luiz Carlos Pereira, and Hermann Haeusler. On what there must be: existence in logic and some related riddles. *Disputatio. Logic, Norms and Ontology.*, 4(34):889–910, 2012.
- Timothy Williamson. *Modal logic as metaphysics*. Oxford University Press, Oxford, 2015.
- Ludwig Wittgenstein. *Tractatus logico-philosophicus*. Routledge, Oxfordshire and New York, 2014.

## LISTA DE PALESTRAS



---

## SEMINÁRIO PERMANENTE LÓGICA NO AVIÃO

(2013-2018)

- 50 Graham Priest (City University of New York): *Evans' Argument and Vague Objects* (maio, 2018);
- 49 Graham Priest (City University of New York): *Everything and Nothing* (maio, 2018);
- 48 Emiliano Boccardi (UNICAMP): *The Delusive Illusion of Passage* (novembro, 2017);
- 47 Emiliano Boccardi (UNICAMP): *Time as Motion* (novembro, 2017);
- 46 Alfredo Roque Freire (UNICAMP): *Existem traduções? Uma crítica às implicações da subdeterminação das traduções em Quine* (outubro, 2017);
- 45 Abílio Rodrigues (UFMG): *Uma abordagem epistêmica para a paraconsistência: a lógica da evidência e verdade* (agosto, 2017);
- 44 Cassiano Terra Rodrigues (PUCSP): *Lógica e quantificação no século XIX* (agosto, 2017);
- 43 Nelson G. Gomes (PPGFIL-UnB/CNPQ): *Negação e contradição* (junho, 2017);
- 42 Daniel Ventura (UFG): *Tipos com Interseção em Computação* (maio, 2017);
- 41 Fernando Ferreira (Universidade de Lisboa): *Ziguezague e a aritmética de Frege* (maio, 2017);
- 40 Edelcio G. de Souza (USP): *O conceito de paradedução* (maio, 2017);
- 39 Fabien Schang (Universidade Estadual de Maringá): *Partition Semantics and its Normal Forms* (maio, 2017);
- 38 Flávio L. C. de Moura (CIC-UnB): *Formal Proofs for Undergraduates* (abril, 2017);
- 37 Hilan Bensusan (FIL-UnB): *A metafísica da contingência* (dezembro, 2016);
- 36 Fabien Schang (Higher School of Economics - Russia): *Introduction to Philosophical Logic - Part 2: Pluralist Pluralism* (dezembro, 2016);
- 35 Fabien Schang (Higher School of Economics - Russia): *Introduction to Philosophical Logic - Part 1: The Ways of Disagreement* (dezembro, 2016);
- 34 Fabien Schang (Higher School of Economics - Russia): *Epistemic Pluralism* (dezembro, 2016);
- 33 Walter Carnielli (CLE, IFCH-UNICAMP): *Sobre deuses, enigmas e o pluralismo lógico* (novembro, 2016);
- 32 Matthias Schirn (University of Munich): *Truth - Frege and Deflationism* (novembro, 2016);
- 31 Mauricio Ayala-Rincón (CIC-MAT/UnB): *Automação da Terminação* (outubro, 2016);
- 30 Hugo L. Mariano (USP): *Identidade entre lógicas* (setembro, 2016);
- 29 André Leclerc (UnB-UFC-CNPq): *Discurso Ficcional e Artefatos Abstratos* (agosto, 2016);
- 28 Yuri Balashov (University of Georgia): *Time, Persistence and temporal experience* (julho, 2016);
- 27 Daniel Durante (UFRN): *Sem divergência lógica não há divergência metafísica e vice-versa* (junho, 2016);

- 
- 26 Evandro L. Gomes (UEM): *Impasses filosóficos acerca da não trivialidade explicativa em contextos teóricos* (maio, 2016);
  - 25 Alfredo Roque Freire (UNICAMP): *Classes e conjuntos* (abril, 2016);
  - 24 Daniele Nantes (MAT-UnB): *Lógica Equacional e Criptografia* (março, 2016);
  - 23 Olavo Leopoldino da Silva Filho (FIS-UnB): *Modelagem de Teorias Científicas e a noção de Verdade Aproximada* (dezembro, 2015)
  - 22 Jairo José da Silva (CNPq/UNESP): *O conceito de completude lógica: Hilbert e Husserl* (outubro, 2015);
  - 21 Fabien Schang (Higher School of Economics - Russia): *Logical disagreements and their galaxies* (setembro, 2015);
  - 20 Edgar Almeida (UNICAMP): *Bases à Verdade de Sentenças Aritméticas* (agosto, 2015);
  - 19 Giorgio Venturi (UNICAMP): *Mathematical explanation and its practice* (junho, 2015);
  - 18 André Porto (UFG): *Mutação semântica e o problema da parada em Wittgenstein* (maio, 2015);
  - 17 Walter Carnielli (UNICAMP): *O que a Lógica deixa de não ser* (abril, 2015);
  - 16 Jean-Yves Béziau (UFRJ/ UC San Diego): *Um círculo quadrado não é uma contradição* (março, 2015);
  - 15 David Rey (Universitat de Barcelona): *Temporal operators, double indexing and expressability* (dezembro, 2014);
  - 14 José Veríssimo Teixeira da Mata (Câmara dos Deputados): *Da Interpretação: aspectos de uma permanência* (junho, 2014);
  - 13 Edelcio G. de Souza (USP): *Sobre estruturas abstratas de consequência* (maio, 2014);
  - 12 Danilo Marcondes (PUCRJ): *Ceticismo e Lógica* (maio, 2014);
  - 11 Décio Krause (UFSC): *A lógica da lógica quântica* (outubro, 2013);
  - 10 Alexandre Madeira (Universidade de Aveiro): *Towards logics hybridization* (outubro, 2013);
  - 9 Rodrigo Bonifácio (CIC-UnB): *Raciocínio sobre programas e cálculo lambda* (julho, 2013);
  - 8 Décio Krause (UFSC): *A teoria de quase-conjuntos* (junho, 2013);
  - 7 Edward Hermann Haeusler (PUC-RJ): *Lógica de Descrição Intuicionista e Formalização de Legislação* (maio, 2013);
  - 6 Claus Akira Horodyski Matsushige (MAT-UnB): *Lógica e Fundamentos da Matemática* (maio, 2013);
  - 5 Diogo Dias (PUC-SP): *Aspectos históricos e filosóficos das lógicas paraconsistentes* (maio, 2013);
  - 4 Claudia Nalon (CIC-UnB): *Resolução para lógicas modais normais* (abril, 2013);
  - 3 Oliver Kutz (Univesität Bremen): *Connecting Logics* (abril, 2013);
  - 2 María Manzano (Universidad de Salamanca): *Logic and Fiction* (abril, 2013);
  - 1 Graham Priest (City University of New York): *Dialetheism: What is it all About?* (abril, 2013).

I.S.B.N. 978-65-00-06187-1