SABERES

REVISTA INTERDISCIPLINAR DE FILOSOFIA E EDUCAÇÃO

NAT@Logic 2015

$\Phi\iota\lambda o\mu\varepsilon\nu\alpha$ II

João Daniel Dantas,
Patrick Terrematte,
Sanderson Molick (Eds.)

Natal-RN
UFRN
2016

SABERES
REVISTA INTERDISCIPLINAR DE FILOSOFIA E EDUCAÇÃO

NAT@Logic 2015

Φιλομενα II

João Daniel Dantas,
Patrick Terrematte,
Sanderson Molick (Eds.)

SABERES
REVISTA INTERDISCIPLINAR DE FILOSOFIA E EDUCAÇÃO
Vol. 1 – N. Esp. $\Phi\iota\lambda o\mu\varepsilon\nu\alpha$ II – Jun.
ISSN 1984-3879

Editors:
João Daniel Dantas,
Patrick Terrematte,
Sanderson Molick.

Sponsorship:

This title can be downloaded at:
http://www.periodicos.ufrn.br/saberes/index

ISSN: 1984-3879
Concept of design: Ulrike Harbort. LaTeX classes: Timm Lichte.

Edited by Patrick Terrematte — XƎLaTeX

Department of Philosophy
UFRN, Natal - RN, Brazil

SABERES
REVISTA INTERDISCIPLINAR DE FILOSOFIA E EDUCAÇÃO
Vol. 1 – N. Esp. $\Phi\iota\lambda o\mu\varepsilon\nu\alpha$ II – Jun.
ISSN 1984-3879

Contents

Introdução Editorial

João Daniel Dantas,
Patrick Terrematte,
Sanderson Molick

SABERES – Edição Especial
$\Phi\iota\lambda o\mu\varepsilon\nu\alpha$ II
(FILOMENA 2)

Criada em 2008, a SABERES: Revista Interdisciplinar de Filosofia e Educação (ISSN 1984-3879) é uma publicação de fluxo contínuo com qualificação B4 pela CAPES, aberta para pesquisadores das áreas de Filosofia e Educação. A publicação constitui um canal de divulgação científica defendendo os princípios de pluralidade e inter-relação das áreas de conhecimento e produção do saber. Vinculada atualmente ao grupo de pesquisa Fundamentos da Educação e Práticas Culturais, e em colaboração com os grupos de pesquisa de Ética e Filosofia Política, e de Lógica, Conhecimento e Ética da UFRN.

A SABERES tem como missão reunir pesquisadores seniores e iniciantes do Brasil e do exterior numa linha editorial que compreenda trabalhos de Filosofia, Filosofia da Educação e Ensino de Filosofia sem qualquer restrição prévia de temática ou de abordagem, buscando reunir acadêmicos de Programas de Pós-Graduação em geral.

A segunda edição do $\Phi\iota\lambda o\mu\varepsilon\nu\alpha$ Workshop (Filosofia, Logica e Metafísica analítica), promovido pelo Grupo de Lógica e Filosofia Formal da UFRN, tem por propósito unir lógicos que trabalham na intersecção entre Lógica e Metafísica, através de aplicações de métodos formais em Filosofia. Lógica, enquanto inicialmente considerada como um ramo da Filosofia, superou seus propósitos originais e encontrou ligações com outras áreas da filosofia, tais como Filosofia da Linguagem, Filosofia da Matemática, Filosofia da Ciência e Filosofia da Mente. Desde o seu desenvolvimento moderno, Lógica se demonstrou ser uma ferramenta poderosa para analisar diferentes teorias filosóficas, assim como seus fundamentos e implicações; além disso, o nascimento e desenvolvimento das Lógicas não-clássicas expandiu seu domínio de aplicação muito além dos sonhos de seus progenitores.

Tópicos de interesse para nosso Worshop incluem, mas não estão limitados a:

- Metafísica Modal
- Referência e Descrições
- Tópicos filosóficos em Lógicas não-clássicas
- Valores de verdade
- Consequência Lógica
- Pluralismo Lógico x Monismo Lógico
- Neutralidade Lógica e Metafísica
- Paradoxos

A primeira edição do FILOMENA, realizado na UFRN em Setembro de 2014, contou com a participação 15 comunicações e uma mesa redonda sobre “The philosophy of contradictions”, com os seguintes participantes: Hitoshi Omori (CUNY), Peter Verdée (KU-Leuven), Marcos Silva (UFC), Chair: João Marcos (UFRN).

Editorial Introduction

João Daniel Dantas,
Patrick Terrematte,
Sanderson Molick

SABERES – Special Issue
$\Phi\iota\lambda o\mu\varepsilon\nu\alpha$ II
(FILOMENA 2)

Founded in 2008, SABERES: Interdisciplinary Journal of Philosophy and Education (ISSN 1984-3879) is a publication of continuous flow holding qualification B4 by CAPES, open to researchers working on the fields of Philosophy and Education. The publication is a space for science communication defending the principles of plurality and interdisciplinarity across different areas of knowledge and knowledge production. Currently linked to the research group Foundations of Education and Cultural Practices, and in collaboration with the research groups of Ethics and Political Philosophy and Logic, Knowledge and Ethics/UFRN.

SABERES's mission is to bring together senior and undergraduate researchers in Brazil and abroad in an editorial line that seeks for works on Philosophy, Philosophy of Education and Philosophy Teaching without any prior restriction to theme or approach.

The second edition of **FILOMENA**, held at UFRN in September 2015, counted with 22 contributed talks. Different topics of the philosophy of logic were discussed and explored within the Workshop. The following edition gathers some of the works presented at the workshop. It has eight papers discussing different topics from philosophy of mind to modal logics and paraconsistent logics.

The $\Phi\iota\lambda o\mu\varepsilon\nu\alpha$ Workshop (Philosophy, Logic and Metaphysics), promoted by the **Group on Logic and Formal Philosophy** from the UFRN, has the purpose of gathering logicians working at the intersection of Logic and Metaphysics, through the application of formal methods in Philosophy. Logic, while initially considered as a branch of Philosophy, has outgrown its original purposes and found connections with other areas of Philosophy, such as Philosophy of Language, Philosophy of Mathematics, Philosophy of Science and Philosophy of Mind. Since its modern development, Logic has proved to be a powerful tool for analyzing different philosophical theories, as well as their foundations and implications; moreover, the birth and development of non-classical logics has expanded its domain of application much beyond the dreams of its progenitors.

Topics of interest for our Workshop include, but are not limited to:

- Modal metaphysics
- Reference and descriptions
- Philosophical topics in non-classical logics
- Truth-values
- Logical consequence
- Logical pluralism x logical monism
- Logic and metaphysical neutrality
- Paradoxes

PARACONSISTENT CONTRADICTION IN CONTEXT

Jonas Rafael Becker Arenhart *

June 28, 2016

Resumo: Dizem que as lógicas paraconsistentes domesticam contradições: em tais lógicas, expressões como α e $\neg\alpha$ não trivializam a teoria. Neste sentido, podemos violar a Lei de Não-Contradição. O principal problema com esta caracterização diz respeito ao fato de que sempre que expressões como α e $\neg\alpha$ são ambas verdadeiras, *elas já não formam mais uma contradição*, mas no melhor dos casos uma *subcontrariedade*. Assim, talvez o maior desafio consista em explicar o que se quer dizer com uma 'contradição paraconsistente', ou seja, por um par de expressões como α e $\neg\alpha$ quando '$\neg$' é uma negação paraconsistente. Sugerimos que há um sentido razoável segundo o qual estas expressões podem ser entendidas, envolvendo a introdução de contextos distintos a partir dos quais α e $\neg\alpha$ são proferidas. Neste sentido, podemos ler estas expressões como formulando uma subcontrariedade and ainda atribuir um sentido intuitivo para 'contradições paraconsistentes'.

Palavras-chave: Lógica paraconsistente. Contradição. Subcontrária. Negação.

Abstract: Paraconsistent logics are said to domesticate contradictions: in such logics, expressions such as α and $\neg\alpha$ do not trivialize the theory. In this sense, we are able to violate the Law of Non-Contradiction. The main problem with this kind of characterization concerns the fact that whenever expressions such as α and $\neg\alpha$ are both true, *they are no longer a contradiction*, but at best a *subcontrariety*. So, perhaps the biggest challenge consists in explaining what is meant by a 'paraconsistent contradiction', that is, by a pair of expressions such as α and $\neg\alpha$ when '$\neg$' is a paraconsistent negation. We suggest that there is a sensible sense in which such expressions may be understood, involving the introduction of distinct contexts from which α and $\neg\alpha$ are uttered. In this sense, we can read those expressions as formalizing subcontrariety and provide for an intuitive meaning for 'paraconsistent contradictions'.

Keywords: Paraconsistent logics. Contradiction. Subcontrary. Negation.

*Professor of Philosophy at the Federal University of Santa Catarina, Florianópolis, Brazil.

Introduction

Paraconsistent logics were designed to deal with contradictions and contradictory statements. To confirm that the aim of paraconsistent logic is related with contradictions it is enough to check the many papers, books and collections published every year that intend to deal with paraconsistency. Once we employ a paraconsistent system of logic, we are told that we do not need to fear contradictions, given that the worst consequence of a contradiction — explosion — does not hold in such logics:

$$\alpha, \neg\alpha \nvdash \beta.$$

That is, in paraconsistent logics contradiction does not imply trivialization of the system (clear examples of such claims is to be found in authors having as different views on the subject as da Costa (COSTA, 1989), (COSTA N. C. A.; BUENO, 2006), and Priest (PRIEST, 2006)).

The most important consequence of such an analysis, it is said, is that one of the most famous laws of traditional logic, the *Law of Non-Contradiction*, is violated: we are allowed to have both α and $\neg\alpha$, true contradictions, as it were. We are also able to deal with contradictory objects, such as those featuring in Meinongian metaphysics, deal with inconsistent data bases, and, perhaps most importantly, keep rationality and sanity while still embracing the contradictions daily life and even our best scientific theories impose on us. The chapter of "Philosophical Motivations and Applications" of paraconsistent logics is indeed a long one.

Perhaps the greatest challenge to such a view of paraconsistency — and it seems to be the only view on paraconsistency available — is to make sense of so-called 'paraconsistent contradictions': what do the expressions α and $\neg\alpha$ mean exactly in paraconsistent logics? How can it be that both are true? Following Beall (BEALL, 2004, p.15), there are at least three distinct approaches to that question:

Weak paraconsistentist: The weak paraconsistentist believes that paraconsistent logics are just interesting mathematical models; useful, but not representing any real possibility. This is a kind of instrumental approach to paraconsistency.

Strong paraconsistentist: this kind of paraconsistentist believes that paraconsistent logics represent real possibilities, even though such possibilities do not obtain in the real world. That is, there are no true contradictions in fact.

Dialetheism: the Dialetheist paraconsistentist believes that some contradictions are true in this world, even if they are not true of concrete observable facts.

As we can see, the three kinds of paraconsistentists diverge on how deep a contradiction reaches. Anyway, and that is the point here, they all agree that paraconsistent logics deal with contradictions. However, there is still a problem of how should we understand this notion of contradiction. As Carnielli and Rodrigues (CARNIELLI; RODRIGUES, 2012, p.2) put it:

> It is a fact that paraconsistent logics permit extending conventional reasoning to make it able to deal with contradictions. However, the nature of the contradictions that are dealt with by paraconsistent logics is still an open issue.

It is curious to find out that contradictions could have distinct natures — in particular that their natures may depend on the systems of logic they appear in —, so that paraconsistent contradictions, in particular, have a nature not so well understood until now. Furthermore, having a nature distinct from classical contradictions, one wonders how paraconsistent contradictions can in fact be the violation of the classical contradiction, the one appearing in the classical law of Non-Contradiction. Now, leaving those oddities aside, our proposal in this paper is to address such an issue by somehow attempting to dissolve part of the problem (although not in a Wittgensteinian fashion). We shall advance the (already widespread) thesis that whenever paraconsistent logics are applied successfully — in philosophical analises of situations apparently involving a contradiction — we are in fact not dealing with a contradiction; in fact, we are dealing with subcontraries, a distinct (weaker) logical opposition. That claim may seem absurd at first, but we hope to make it reasonable by calling the reader's attention to the fact that whenever we violate the explosion law, by having a counter-model to the Explosion inference $\alpha, \neg\alpha \vdash \beta$, we must face the fact that the pair $\alpha, \neg\alpha$ represents only subcontraries.[1] So, in the end of the day, our proposal (we hope) seems to bring together the best of both worlds: we avoid the need to deal with contradictions (and to understand their nature in paraconsistent logics) and still make sure that applications of paraconsistent logic do reflect the phenomena addressed by them.

The plan of this paper is as follows. In section 2, to make the paper self-contained we review some of the terminology employed and present some arguments favoring a semantic treatment of the oppositions. In section 3 we briefly present some of the most well-known arguments to the fact that paraconsistent logics deal with subcontrariety, presenting the challenge to paraconsistent logics: making sense of paraconsistent contradictions, *i.e.*, expressions of the form α and $\neg\alpha$. In section 4 we present our proposal of informal interpretation of paraconsistent negation, which turn a paraconsistent contradiction into at best a relation of subcontrariety. We argue that this move somehow dissolves the problem of contradiction in paraconsistent logics. We conclude in section 5, and following Béziau (BÉZIAU, 2014) and (BÉZIAU, 2015), argue that we would be better not talking about contradictions in paraconsistent logics.

Contradiction and Other Oppositions

We shall now introduce some of the terminology that sets the stage for the problem we are dealing with. It is very common to define a contradiction in purely syntactical terms: two expressions are contradictory if they are of the form α and $\neg\alpha$, where $\neg$ is the negation sign of the language. There is, however, some good reasons for us not to follow this approach here. The first reason is that by accepting this definition of contradictory propositions we would be able to have an easy answer, an answer by definition, to the problem we are dealing with: any language with an operator that is supposed to play the role of a negation sign would automatically be able to represent contradictions. Now, there is good grounds to call such an attempt as begging the question, given that what we are investigating now is precisely the relation between negation and contradiction; for those pursuing this kind of answer, anyway, the problem of the 'nature of the contradiction' being represented would still be left hanging in the air.

Related with this first difficulty, there is another one that appears immediately after careful consideration of the previous paragraph: defining contradiction with the use of negation throws under the carpet too many important problems. For instance, it assumes that every sign that is supposed to act as a negation is really a negation, by the simple magical fact that we have stipulated it. As we know, that move goes just too fast; there is a great deal of philosophical discussion on which are the properties an operator *must* have to be legitimately called *negation*, and there is also just as much discussion about which properties a negation sign is allowed to

[1] As we shall see soon, that point is almost universally conceded by both friends and foes of paraconsistency; see Slater (SLATER, 1995) and Béziau (BÉZIAU, 2006).

lose while *still being a negation*. Considering an extreme case, Slater (SLATER, 1995), for one, does not consider paraconsistent negations as negations, so that assuming that we have a negation sign that always characterizes contradictions would amount to simply ignore the sceptic instead of arguing with him. In the end, conflating the concepts of negation and contradiction assumes to much to begin with (it takes those disputes on what negation is as irrelevant or as already solved). When there is so much at stake one cannot simply jump over such problems and assume by stipulation that a sign with a given number of properties is a negation.

There is a third main reason not to define contradiction syntactically, with the help of negation: traditionally, *contradiction is a semantic notion*. Given that our claim in this paper is that the syntactic expression in some cases does not capture the semantic content of a contradiction, defining contradiction in syntactic terms begs the question against any such discussion of 'expressive adequacy'. Also, there are cases in which we are clearly facing a contradiction but no negation is in front of one of the expressions. Consider the Aristotelian pair "Every men is mortal", as a contradictory of "Some men are not mortal". Of course, negation appears inside the second proposition, but it is not the case that we have here expressions of the form α and $\neg\alpha$, while it is at the same time clear that we are facing a contradiction, if anything is a contradiction. For another example, consider an electron that has just been measured for spin in a given direction. In this case, the propositions 'the spin is up' and 'the spin is down' are contradictory: both cannot be true (no electron can be measured both spin up and down at the same time), and both cannot be false, at least one of them is true (the measurement outcomes are comprised only between those two results). Again, we have no negation, but we still have a contradiction.[2]

But couldn't we adopt an inferentialist semantics and hold that the meaning of a logical constant such as negation is given by its introduction and elimination rules in a natural deduction system (under some reasonable conditions), for instance? We could, indeed. However, notice that distinct logical systems (classical, paraconsistent, intuitionist) have distinct rules for such connectives. That would imply that the meanings of negation are different in distinct systems, from which it follows that expressions of the form α and $\neg\alpha$ would also have distinct meanings. The best we could do in this case would be to speak of '*classical* contradiction', '*intuitionist* contradiction', '*paraconsistent* contradiction', and so on. That multiplication of contradictions would only contribute to the conclusion that when a paraconsistent logician says that the Law of Non-Contradiction does not hold, he is not in fact . . ., well, contradicting, his fellow classical logician. To use an expression that appeared earlier, those formulas have distinct 'natures' in distinct systems. So, again, that is not the best option for our purposes.

It is well known that some paraconsistent logics were originally introduced as axiomatic formal systems; only later a semantics was developed. Anyway, even those systems were said to violate the Law of Non-Contradiction, mainly due to the fact that expressions such as $\alpha \wedge \neg\alpha$ were seen as not trivializing a theory in every case. The main point that we need to address is whether that expression is really a contradiction. To define contradictions semantically we employ the terminology of the square of opposition (following, for instance, Béziau (BÉZIAU, 2003)):

[2] Obviously, that same reasoning does not carry to the case of an electron in a superposition of the states 'spin up' and 'spin down' in a given direction, when no measurement is being carried out. For more discussions relating quantum mechanics and oppositions, in particular when superpositions are taken into account, see Arenhart and Krause (ARENHART; KRAUSE, 2016).

Contradiction: Propositions α and β are *contradictory* when α and β cannot both be true and cannot both be false.

Contrary: Propositions α and β are *contrary* when both cannot be true, but both can be false.

Subcontrariety: Propositions α and β are *subcontrary* when both can be true, but both cannot be false.

Again, notice that the definition is not in syntactical terms of a negation operator, but in semantic terms of truth values. Of course, definitions are judged not from the point of view of truth and falsity (*i.e.*, definitions are not true or false), but rather from their faithfulness, fruitfulness, utility and so on. We hope that the reader agree that the semantic definition of contradiction and of the other oppositions do satisfy those conditions.

It is in terms of those definitions of contradiction and subcontrariety that we shall explore the problem of understanding what exactly a pair of expressions of the kind α and $\neg\alpha$ means.

Paraconsistent Contradictions

Now, we can relate paraconsistent negations and the semantic opposition of subcontrariety in the following terms (following Béziau (BÉZIAU, 2006) and Dutilh Novaes (NOVAES, 2008)). Assume that we want to violate the Explosion rule

$$\alpha, \neg\alpha \vdash \beta.$$

What are we supposed to do? We must present a model (valuation, situation, world...) in which both α and $\neg\alpha$ are true, while β is false. However, when that is done and both α and $\neg\alpha$ are true we no longer have a contradiction. We have only the possibility of a subcontrariety. In fact, in general it is accepted that when α is false, then $\neg\alpha$ is true; this grants that we always have that at least one of the members of the pair is true (sometimes only one is true, sometimes both are true). This is enough to get a subcontrariety. In other systems, comprising paranormal logics, we have the possibility that both α and $\neg\alpha$ be false (see Béziau (BÉZIAU, 2015)). In these cases, however, we must also have both true when explosion fails. So, paraconsistent negations represent at best subcontrariety, but they do not represent contradictions.

Some problems now arise: how are we supposed to solve problems (philosophical, practical) involving contradictions by applying an operator that does not represent contradictions? Is that a legitimate move? Even though we shall not discuss those practical issues here, they are relevant for a proper understanding of the scope and limits of paraconsistency — and are related with our discussion to follow (for further discussions on these issues, see Arenhart (ARENHART, 2015)). Our concern in what follows will focus rather on what Dutilh Novaes (NOVAES, 2008) called "the challenge for the paraconsistentist", that is, explaining how can we distinguish situations in which it is legitimate to have both α and $\neg\alpha$ without explosion, from those cases in which α and $\neg\alpha$ cause explosion.

Challenge: How to distinguish cases in which contradictions are explosive — classical contradictions — from cases in which contradictions are not explosive — paraconsistent contradictions?

In other words, we shall spell the problem by employing the terminology we introduced in the previous section: how to distinguish cases which may be dealt with by a subcontrariety forming operator from those cases in which a real case of a contradiction is involved? As it is implicit in the challenge advanced by Dutilh Novaes, there are cases in which not even a paraconsistent logic is enough to avoid explosion.

Notice that this attitude towards contradictions is distinct from Priest's (PRIEST, 2004) claim that even though *some* contradictions may be true, it is not the case that *every* contradiction may be true. Priest's claim is based on the advice that sometimes, even though we could accept true contradictions, other considerations (such as simplicity and fertility of an alternative explanation to the phenomenon under consideration) may lead us to avoid such a path and look for a consistent treatment of the problem. Dutilh Novaes' claim is distinct: some contradictions could not be dealt with by a paraconsistent logic even if we wanted to. They are not amenable to treatment by a paraconsistent logic, or, as we shall say, they are not 'paraconsistent apt'. This is enough to express a deep disagreement on the nature of paraconsistent logic: in one case (Priest's), it seems implicit that a paraconsistent logic can be applied with no limitations; the only limit is imposed on us not from the nature of contradictions and paraconsistency, but from extraneous considerations such as the call of pragmatic virtues. We shall restrict ourselves to the problem raised by Dutilh Novaes. Here the problem is to distinguish where there are legitimate cases of application of paraconsistent logics: it is not possible to tame every contradiction with paraconsistent logics.

We believe that the proper answer to the question as to what exactly counts as a 'paraconsistent contradiction', to address the worries raised in the paper by Dutilh Novaes, may be found in the literature on paraconsistent logics. In fact, our strategy to address the issue will be to present and analyze here some of the cases in which paraconsistent logics are called to solve an apparent case of contradiction and to argue that there is no contradiction involved, but only a subcontrariety. What makes those cases 'paraconsistent apt' is precisely that: those are cases in which we may find a subcontrariety where an apparent contradiction is present. Cases that are 'paraconsistent inapt', that is, cases where contradictions lead to trivialization (*i.e.*, where there is really a contradiction, and not a subcontrary) are resistent to such an interpretation.

How do apparent contradictions become subcontrariety? As we shall see, statements that are usually taken as contradictories may be seen as subcontraries by the introduction of a kind of context before a statement and also before its negation. Taking the context into account transforms contradictions into subcontrarieties, and so we have a non-explosive situation.

But now the question may be: what is a context? As we shall see by the examples presented soon, they may be propositional attitudes, quantifiers, modal operators, points of views, and much more. Their aim is to transform a pair of contradictory sentences into a pair of subcontrary utterances.

Contradictions in Context

In this section we address directly the problem raised by the challenge presented in the previous section.

The first example we shall discuss comes from Jean-Yves Béziau, which in his recent paper (BÉZIAU, 2014) presents arguments to the effect that paraconsistent logics deal with subcontrariety, and not with contradictories. In fact, he was one of the first paraconsistent logicians to go on and advance the suggestion that we should stop speaking about contradictions, keeping only at best subcontrariety. By discussing this case here we are not implying that Béziau is not aware that paraconsistent logics do not deal with real cases of contradiction; rather the other way around (although Béziau still speaks about ‘contradictory viewpoints’). This example is being discussed because we think of it as paradigmatic case of application of paraconsistency, illustrating the strategy of introducing contexts we mentioned before.

What is more relevant for us is his case study of a paraconsistent system which incorporates the idea of contradictory points of view, and this illustrates perfectly the kind of strategy of introduction of a context to render a contradiction into a subcontrariety. Béziau considers an aspect of the quantum mechanical description of reality: the paradoxical wave-particle duality. As it is well-known, in the context of the famous two-slit experiment, for example, quantum objects behave sometimes as waves, sometimes as particles.[3] In this kind of experiment, a coherent light source is sent through a plate having two parallel slits; there is a screen after the slits that registers the arrival of the items[4] going through the slits. When the light beam is sent and both slits are open, the wave nature of light makes the light that passed through both slits interfere, so that a pattern is generated on the screen after the plate; a pattern, indeed, that would be expected of the behavior of a wave (see the picture in (GHIRARDI, 2005, p.50), for instance). On the other hand, when only one slit is open, the pattern created by the particles on the screen after the slit is just what would be expected from particles going through the single slit, no interference in this case.

Bohr and the so-called *Copenhagen Interpretation* of Quantum Mechanics interpreted the situation through the famous *Complementarity Principle*: particles behave both as waves and as particles, but never in the same context.[5] It is the experimental apparatus that determines what kind of observation will result, but we never experiment both aspects; they are complementary and provide for a complete — even if both are excluding — description of reality.

From a logical point of view, the idea of distinct points of views can be easily formalized, as Béziau (BÉZIAU, 2014) has shown, by employing the tools of modal logic and possible world semantics (which are then rebaptized as viewpoints) with a universal accessibility relation. Let us represent by P the statement that O is a particle and by Q the statement that O is a wave. Consider a point of view w_1, according to which O behaves as a particle; then, according to w_1, it is true that O is a particle ($v(P,w_1) = V$), and false that it is a wave ($v(Q,w_1) = F$). Now, consider another point of view w_2, according to which O is a wave. Then, according to w_2 it is true that O is a wave ($v(q,w_2) = V$, and it is false that O is a particle ($v(P,w_2) = F$). The situation may be put schematically thus:

Viewed from w_1**:** $v(P,w_1) = V$, $v(Q,w_1) = F$.

Viewed from w_2**:** $v(Q,w_2) = V$, $v(P,w_2) = F$.

Assuming an universal accessibility relation between worlds, Béziau defines a negation as follows:

$$v(\sim p, w_i) = F \text{ in point of view } w_i \text{ if } v(p,w) = V \text{ for every point of view } w.$$

That is, the negation of a proposition p is false in a point of view w_i if p is true in every point of view associated with w_i. Let us put that definition to work in the example of the particle

[3] We shall not present all the details here; for further discussions and details, see for instance Ghirardi (GHIRARDI, 2005).

[4] We use the term ‘items’ here because it seems neutral as to whether those items are particles or waves or objects of any kinds. Given that the problem surrounds precisely the status of those entities, a neutral word seems advisable.

[5] Again, there is much controversy about the precise formulation of the Principle and its intended meaning, but we refer the reader to (GHIRARDI, 2005) for further details and references.

O. In our example, $v(P,w_1) = V$. Given that there is an associated point of view w_2 such that P is false in that point of view ($v(P,w_2) = F$), we have that $v(\sim P,w_1) = V$. That is, both P and $\sim P$ are true at w_1. However, that does not mean that we have a contradiction: this is a case of subcontrariety; in fact, '$\sim$' is a paraconsistent negation. Furthermore, we do not have something O being both a wave and a particle. It is the peculiar way that the paraconsistent negation is introduced that makes the points of view disappear, and put in the same level, let us say, P and $\sim P$, which, on its turn, leads us to think that we are facing a contradiction.

Having said that, let us discuss how it illustrates the general thesis that paraconsistency involves the introduction of a context. From the brief presentation we have described before, it is clear that in the language for the system described, with the semantics of point of view, we can have both P and $\sim P$ true. However, the intuitive reading of this situation does not allow us to read it as a contradiction; we do not read P and $\sim P$ as stating that 'O is a particle' and 'O is not a particle'. As the application to particles illustrates, what we really have when the paraconsistent logic applies is distinct points of view from which a proposition is judged. In fact, P is true means that *from a given point of view* w_i the proposition is true. $\sim P$ is true in w_i means that *from a distinct point of view* the proposition P is false. This is hardly surprising, given how the construction that was made. However, as we shall argue, it is this kind of introduction of a hidden context — in this case a 'point of view' — that makes sense of paraconsistent contradictions.

This is also the same strategy behind the construction of discussive logics (see the details in (COSTA N. C. A.; BUENO, 2006, pp.844-49)). By beginning with a model of $S5$ and allowing that a proposition is true in discussive logic whenever it is true in one world of the model, we are allowing that, taken collectively, propositions p and $\neg p$ may be true. In fact, they are never true *at the same world*, but it only happens that p is true in one world $_i$, while $\neg p$ is true at another world w_j. Given that worlds represent participants of a discussion group, the pair p and $\neg p$ may both be true, that is, it is a case of a subcontrariety again. Also, its intuitive interpretation introduces the hidden context: "participant w_i states…", and "participant w_j states ….

What is common to both examples and to others we shall present soon is the fact that from the point of view of the paraconsistent language in which propositions are regimented, we have a proposition p and its negation $\neg p$. However, as we have discussed in the previous section (and see also (NOVAES, 2008)), a contradiction needs not be characterized by negation. It is in fact a semantic notion. Now, the fact that we are in face of a subcontrariety appears only when we shift back to informal language, when a context that disappears in the formal language needs to be reintroduced to get the informal meaning of the propositions.

The same principle may be found in discussions of application of paraconsistent logics to other group situations, as for instance a disagreement of a group of referees (see (CARNIELLI; RODRIGUES, 2012)) or the construction of expert systems or knowledge bases of a given domain D. In the last case (the reasoning applies also to the case of referees), we may (and probably will) find experts disagreeing about particular cases. Consider the case of doctors evaluating the symptoms of a given patient (this example is taken from (COSTA N. C. A.; BUENO, 2006, p.865)):

> … given the same observable symptoms, doctor d_1 may believe that the patient has a virus infection, doctor d_2 may conclude that the patient has an allergic reaction, while doctor d_3 may say that the patient either has a viral infection or an allergy, but not both. It is clear that if we had used the opinions of these three doctors in our knowledge base, we would be led into an inconsistency.

In this case, the data base would simply "ignore" the fact that the opinions come from distinct sources. That is, what we really have is, for instance, that "According to doctor d_1, the patient has a virus infection", and, let us say, "According to doctor d_2, the patient has no virus infection". Again, as far as that is all that we have, this is a case of a subcontrariety, not a case of contradiction. The patient, by herself, does not 'have and not have' a virus infection (or allergic reaction). So, the intuitive meaning of the propositions bring them back to the context in which they are subcontraries. That is, from the syntactical form p and $\neg p$ in the formalization of the judgements constituting the data base we may forget that we are dealing with subcontrariety. However, the intuitive meaning re-introduces the contexts and keeps sanity. The contexts, in this case, are:

"According to doctor d_1 the patient has a virus infection."

"According to doctor d_2 the patient has not a virus infection."

Another case corroborating the general thesis of this section concerns partial truth and its applications (see da Costa and French (COSTA N. C. A., 2003), da Costa, Krause and Bueno (COSTA N. C. A.; BUENO, 2006)). Let us summarize the general idea. Consider a domain of knowledge Δ and our scientific theorizing about it. We begin by forming a set D consisting of the observable entities in Δ, enlarged later with unobservable entities to account for observable behavior. Given a particular binary relation R holding between the elements of D, we do not always know whether $\langle d_i, d_j \rangle \in R$ or $\langle d_i, d_j \rangle \notin R$. In actual science, it may reasonable be expected that we don't know which is the case. Partial relations account for that and make for a more realistic description of science. Besides the pairs $\langle d_i, d_j \rangle$ which are in R and those which are not, partial relations are characterized by the ordered pairs of which we are ignorant about whether they do belong to R or not. A partial structure is a structure $\mathfrak{A} = \langle D, R_i \rangle_{i \in I}$, where R_i is a family of partial relations. Now, sometimes we can extend a partial structure to a total structure by putting, for each relation of the structure, the ordered pairs $\langle d_i, d_j \rangle$ of whose statuses we are ignorant either in R or definitively out of R. This constitutes, as we mentioned, an extension of a partial structure $\mathfrak{A}$ into a total structure $\mathfrak{B}$. We can now evaluate for the (Tarskian) truth of propositions of the language in $\mathfrak{B}$, given that it is a Tarskian structure. We say that a sentence α is partially true in $\mathfrak{A}$ when there is a total structure $\mathfrak{B}$ in which α is true in the Tarski style.[6]

Now, it is easy to see that both α and $\neg\alpha$ may be partially true in some cases: suffices to have a partial structure $\mathfrak{A}$ such that there are two extensions of it, one $\mathfrak{B}$ in which α is true, and another, $\mathfrak{C}$ in which $\neg\alpha$ is true. So, a contradiction may be partially true, at least? Is that a contradiction? Not really. As the reader may have noticed, α and $\neg\alpha$ are not true *in the same structures*, but rather in total structures $\mathfrak{B}$ and $\mathfrak{C}$, respectively, which extend $\mathfrak{A}$. There is again the introduction of a context that makes a subcontrariety appear, and when the intuitive meaning is sought, we find that distinct contexts make a proposition and its negation partially true.

We could go on and present further examples, but we believe that those are enough to illustrate the general thesis (see also the discussions in Arenhart (ARENHART; KRAUSE, 2016)). When a paraconsistent logic is called to deal with a contradiction, what we are really facing is a subcontrariety. The syntactical form of α and $\neg\alpha$ (when '$\neg$' is a paraconsistent negation) leads us to think that we are facing a contradiction, but we should recall that a contradiction is not defined in terms of negation and that there are the hidden contexts playing their role.

This provides an answer to the challenge present by Dutilh Novaes we mentioned before. Recall that the challenge consists in specifying when we may employ a paraconsistent logic

[6] There are much more details and technicalities in the definition of partial truth; however, for our purposes such a brief account suffices. The reader may check (COSTA N. C. A., 2003) also (COSTA N. C. A.; BUENO, 2006, sec.4.3).

to deal with a contradiction and when the contradiction is just too strong to be dealt with by employing such logics. The answer that arises from the above considerations puts many facts already established in the literature together to tailor an answer to the question. Let us check how it works.

Recall that a paraconsistent negation, in order to be paraconsistent, will have to be at most a subcontrary forming operator. That is, as Béziau (BÉZIAU, 2006, p.23) has put it: when a negation sign forms contradictions it is not paraconsistent, and when it is paraconsistent, it does not form contradictions. So, when we speak about 'paraconsistent contradictions' we are speaking about cases of subcontrarieties (at best). Paraconsistent logics are applicable whenever we have a situation in which an apparent contradiction may be seen as a subcontrariety by the appeal to the appropriate contexts. That is, we must find what are the (almost always) hidden contexts that make the conflicting propositions 'paraconsistent able', *i.e.*, able to be such that a proposition and its negation are both true. As we have seen, such contexts may include modal operators, points of view, propositional attitudes and much more. When such contexts are available we are authorized to apply paraconsistent logics. In a nutshell, *when the contradiction is not really a contradiction, but rather acts as a subcontrariety, there paraconsistent logics are applicable*. The formal apparatus of such logics are best suited to deal with such situations.

As a final problem, let us briefly the question of whether there could be a limit to this kind of "context insertion" strategy. Does every contradiction allow for the introduction of a hidden context, so that paraconsistency applies to every apparent contradiction? It does not seem so. Let us consider the reasons. In a recent debate, Beall (BEALL, 2011) and Eldridge-Smith (ELDRIDGE-SMITH, 2012) illustrate such a situation in which it seems doubtful whether one can reasonably introduce a context. Consider Pinocchio stating the sentence "My nose will grow". According to the story, Pinocchio's nose grows if and only if he tells a lie. So, we are facing a paradox, and in the end Pinocchio's nose will grow and not grow. Beall (BEALL, 2011) does not think that this is a case of a real contradiction. What is really at stake, according to him, is that we have hidden contexts in those statements: "According to the story, Pinocchio's nose will grow", and "According to the story, Pinocchio's nose will not grow". There is not a real contradiction.

So far, so good; this "in the story" introduction illustrates the use of contexts to generate situations in which paraconsistent logics apply, where contradictions become subcontraries. However, Eldridge-Smith (ELDRIDGE-SMITH, 2012) reasonably asks for what makes the Pinocchio case liable to such a treatment inside contexts, while by the lights of some paraconsistentists there are other cases, such as those involving the Liar Paradox, that are not liable to such treatments. That is, as Beall seems to believe, the Liar Paradox seems to involve a real case contradiction, but the Pinocchio case seems to be rendered innocuous by the introduction of contexts. What is the difference? As Eldridge-Smith (ELDRIDGE-SMITH, 2012) puts, the difference seems to rely on metaphysical stipulation, which on their turn must be argued for, not merely stipulated. This bears directly on the question of how to distinguish between paraconsistent apt 'contradictions' and paraconsistent inapt contradictions. The question then is: how could one provide for a principled reasoning to distinguish such cases where a contradiction legitimately appears from those in which a context may be introduced to avoid contradictions?[7]

There may be no principled way to distinguish between the relevant contexts, but an example may help us illustrate what is going on. As we mentioned throughout the section, paraconsistent negation seems apt when there is a contexts inside which it is introduced. The problem arises

[7] Notice that Beall and Eldridge-Smith do not put the problem in those terms, however.

when the negation can somehow escape the context, that is, the real contradictions appear when from a pair Cp and $C\neg p$ we pass on to a pair Cp and $\neg Cp$, where C represents a context and p is a proposition. Let us follow the usual terminology and call "Exclusion" the property a context have when from $C\neg p$ we can legitimately infer $\neg Cp$. It seems exclusion is relevant to answer the question about when a context is able to allow a paraconsistent contradiction.

Now, as we have mentioned, this is not to be taken as a categorical answer, but rather as a first step in this direction. The issue may be clearly illustrated by a discussion advanced by da Costa and French (COSTA N. C. A., 2003, pp.98-100). The main problem they are considering here is whether belief in a contradiction (defined syntactically) would entail contradictory beliefs; that is, by using **B** to be the belief operator, the question is whether $\mathbf{B}(p \wedge \neg p)$ entails somehow $\mathbf{B}p \wedge \neg \mathbf{B}p$. In the first case, notice, the contradiction is confined to the scope of the operator, it is an "internal contradiction", while in the second case, the contradiction spreads outside of the operator, it is an "external contradiction", and we have contradictory beliefs.

The crucial step in going from internal contradictions to external ones (that is, reaching out of the context) concerns application of the Exclusion property:

1. $\mathbf{B}(p \wedge \neg p)$ (assumption)
2. $\mathbf{B}p \wedge \mathbf{B}\neg p$ (distribution of **B**)
3. $\mathbf{B}p \wedge \neg \mathbf{B}p$ (from 2, with Exclusion)

So, the main problem, at least in the case we are considering, seems to grant that we may prevent that the belief operator indeed obey the Exclusion property. According to the diagnosis by da Costa and French (COSTA N. C. A., 2003, p.99), the main problem comes from the association of "belief" with "believing that is true". By assuming that "belief that p" ($\mathbf{B}p$) is synonymous with "belief that p is true" ($\mathbf{BT}p$), the authors claim that we obtain the Exclusive property for belief (see again da Costa and French (COSTA N. C. A., 2003, p.99)):

1. $\mathbf{BT}\neg p \rightarrow \mathbf{B}\neg \mathbf{T}p$ (exclusion for truth)
2. $\mathbf{B}\neg \mathbf{T}p \rightarrow \neg \mathbf{BT}p$
3. $\mathbf{B}\neg p \rightarrow \neg \mathbf{B}p$ (dropping **T** from 2)

Obviously, the problem here concerns step 2. If that is an instance of Exclusion for belief, then the claim by da Costa and French according to which belief results exclusive because of its association with truth is ungrounded. However, they seem to think that that step is justified mainly because of properties of truth:

> [...] the exclusive nature of beliefs seems to follow from that of truth, together with the standard understanding of "belief that p". We might say that there is nothing inherent in belief itself that causes it to be exclusive; it's all in the propositions and whether they are regarded as true or not. (da Costa and French (COSTA N. C. A., 2003, p.99))

Now, even though that conclusion is not obvious from the derivation offered (step 2 seems either to require exclusion for beliefs or else to be mysterious), we may follow the general suggestion that what prevents a paraconsistent context involving belief from becoming a classical context is the fact that one avoids that belief is exclusive, that is, that a negation may become external to a context. Shifting to weaker notions of truth, such as quasi-truth, may allow one to block such inferences and keep belief without exclusion. In this sense, a scientist, for instance, may belief in a contradiction (*i.e.* $\mathbf{B}(p \wedge \neg p)$, the contradiction is inside a context) without having contradictory beliefs (*i.e.* $\mathbf{B}p \wedge \neg\mathbf{B}p$, the contradiction is outside the context).

So, when we reason about our beliefs, $\neg(\mathbf{B}p \wedge \neg\mathbf{B}p)$ should hold, while an appropriate doxastic logic should also allow for $\mathbf{B}(p \wedge \neg p)$. As da Costa and French put it: "when we reason about our own beliefs, our "external" logic is classical, whereas our "internal" logic is paraconsistent" (da Costa and French (COSTA N. C. A., 2003, p.100)). That is, when contradictions are confined to the context of belief, we reason with paraconsistent logics, allowing for contradictions inside the context. When the contradiction spreads out, we are in the field of classical logic, and that case must be avoided.

This illustrates rather well our suggestion that paraconsistent negations are always hiding a context, a context which allows the propositions involved to become subcontraries. When we get out of the context and the propositions become contradictions, then, it is classical logic that is operating.

Final Remarks

If the arguments presented in the previous sections are correct, there is a way in which we can employ tools already available in the literature in order to address the challenge to the paraconsistentist: 'paraconsistent contradictions' are cases of subcontrariety, which may be seen as not really contradictory when we evaluate the statements by taking into account the contexts of utterances. That idea, as we mentioned earlier, helps us making sense of paraconsistent applications while keeping it consistent with well-known facts about paraconsistent negations (it generates subcontraries, and not contradictories).

What the suggestion does not solve is the problem of contradictions. In fact, contradictions are still to be avoided, they do trivialize a system. So, the usual contention that paraconsistent logics help us violate the Law of Non-Contradiction will have to be revised. The Law of Non-Contradiction appears to be violated in paraconsistent logics due to syntactical similarities of formulas in classical and in paraconsistent logics. However, a 'paraconsistent contradiction' does not seem to be a contradiction in the sense that Aristotle condemned. Those are rather distinct things.

Bearing that in mind, some of the *Philosophical Applications* of paraconsistent logics will also have to be revised and a humble attitude to be assumed. It is said that paraconsistent logics deal with the realm of contradictory objects of Meinongian ontology, that it allows us to have Russell's set understood, and that it allows us to be rational in the presence of contradictions. A more realistic attitude towards such applications seems to be advised. As Béziau (BÉZIAU, 2014), (BÉZIAU, 2015) remarked, it would be closer to the facts if we stopped speaking about contradictions when dealing with paraconsistent logics and kept only subcontrariety. That is much less spectacular, but is much closer to the truth.

References

ARENHART, J. R. B. Liberating paraconsistency from contradiction. *Logica Universalis*, v. 9, p. 523–544, 2015.

ARENHART, J. R. B.; KRAUSE, D. Contradiction, quantum mechanics, and the square of opposition. *Logique et Analyse*, Forthcoming, 2016.

BEALL, J. C. Introduction: at the intersection of truth and falsity. In: PRIEST, J. B. G.; ARMOUR-GARB, B. (Ed.). *The Law of Non-Contradiction: New Philosophical Essays*. Oxford: Clarendon Press, 2004. p. 1–19.

BEALL, J. C. Dialetheists against pinocchio. *Analysis*, v. 71(4), p. 689–691, 2011.

BÉZIAU, J.-Y. New light on the square of oppositions and its nameless corners. *Logical Investigations*, v. 10, p. 218–232, 2003.

BÉZIAU, J.-Y. Paraconsistent logics! a reply to slater. *Sorites*, v. 17, p. 17–25, 2006.

BÉZIAU, J.-Y. Paraconsistent logic and contradictory viewpoints. *Revista Brasileira de Filosofia*, v. 241, 2014.

BÉZIAU, J.-Y. Round squares are no contradictions (tutorial on negation, contradiction, and opposition). In: BéZIAU, M. C. J.-Y.; DUTTA, S. (Ed.). *New Directions in Paraconsistent Logic*. New Delhi: Springer, 2015.

CARNIELLI, W.; RODRIGUES, A. What contradictions say (and what they say not). *Cle e-prints*, v. 12(2), 2012.

COSTA, N. C. A. da. The philosophical import of paraconsistent logics. *CLE Manuscripts*, 1989.

COSTA N. C. A., F. S. da. *Science and Partial Truth: a Unitary Approach to Models and Scientific Reasoning*. Oxford: Oxford Un. Press, 2003.

COSTA N. C. A., K. D. da; BUENO, O. Paraconcistent logic and paraconsistency. In: JACQUETTE, D. (Ed.). *Handbook of the Philosophy of Science. Volume 5: Philosophy of Logic*. Amsterdam: Elsevier, 2006. p. 791–911.

ELDRIDGE-SMITH, P. Pinocchio beards the barber. *Analysis*, v. 72(4), p. 749–752, 2012.

GHIRARDI, G. *Sneaking a Look at God's Cards: Unraveling the Mysteries of Quantum Mechanics*. Princeton: Princeton Un. Press, 2005.

NOVAES, C. D. Contradiction: the real philosophical challenge for paraconsistent logic. In: BéZIAU W. CARNIELLI, D. G. J.-Y. (Ed.). *Handbook of Paraconsistency*. Amsterdam: Elsevier, 2008. p. 465–480.

PRIEST, G. What's so bad about contradictions? In: PRIEST, J. B. G.; ARMOUR-GARB, B. (Ed.). *The Law of Non-Contradiction: New Philosophical Essays*. Oxford: Clarendon Press, 2004. p. 23–38.

PRIEST, G. *Doubt Truth to be a Liar*. Oxford: Oxford Un. Press, 2006.

SLATER, B. H. Paraconsistent logics? *Journal of Philosophical Logic*, v. 24, p. 451–454, 1995.

CONDITIONAL ANTINOMIES

Claudio Pizzi *

Abstract: After a short premise about some paradoxical features of counterfactual reasoning, the paper tries to identify the analogues of the so-called Card Paradox and of the Liar antinomy by using a language containing the conditional operator symbolized by $>$. A first proposal is taken into account but dismissed since the resulting Liar is equivalent to the statement "the present statement is necessarily false", which makes the corner operator indistinguishable from strict implication. A second proposed solution consists in paraphrasing every statement *A* into the conditional "I would lie if I said not-A", where the conditional has the properties of the classical conditional operator. The "Epimenides" and "Truth-Teller" variants of the paradox are also examined in the last section.

Keywords: Antinomy. Paradox. Liar. Epimenides. Counterfactuals. Conditionals.

*Emeritus Siena University. Born in Milan (Italy) in 1944. Main research interests: Tense Logic, Conditional Logic, Philosophy of Causality, Abductive Reasoning.

§1. Conditionals have been a source of paradoxes since the beginning of contemporary logic, as it is shown by the so-called paradoxes of material and strict implication. However, their more paradoxical features came out especially after the II World War, along with the first attempts to determine the logical properties of counterfactual conditionals[1]. It is in itself paradoxical, to begin with, that a contrary-to-fact antecedent has normally two legitimate consequents which are incompatible. For instance, under standard presuppositions about Apollo and human kind, both the following counterfactuals appear to be justified:

(i) If Apollo were a man he would be mortal

(ii) If Apollo were a man he would be the instance of an immortal man

---

The paradox is here solved by the fact that one of the two conditionals, i.e. (ii), is "less intuitive" than the other. This asymmetry may be explained in terms of the possible world semantics which David Lewis associated to conditionals: the worlds in which Apollo is a man and is mortal are more similar to the actual world than the worlds in which Apollo is a man and is immortal. This explains why (ii) should be rejected and (i) should be accepted.

A different approach to conditionals is the Chishom-Goodman-Reichenbach one, according to which the truth of a conditional depends on a consequential nexus between the clauses. We may propose a definition of a consequential nexus in the following way.

We observe that (i) turns out to be more acceptable than (ii) since the consequent of (ii) implies the negation of a law of nature, i.e the removal of something which has a high information content, while in (i) the consequent implies the removal of a singular fact, whose information content is lower that the one of any law of nature. The latter consideration allows defining the consequential nexus between A and B, in a conditional of form $A > B$, as something which subsists when $A > B$ is more acceptable, in the defined sense, than $A > \neg B$. On the ground of such definition of truth, the truth of $A > B$ implies that $A > \neg B$ is false, or alternatively that $\neg(A > \neg B)$ is true. The rule $A > B \vdash \neg(A > \neg B)$ is sometimes called "Boethius' rule".

The main problem yielded by both the consequentialist and the Lewis' semantics for conditionals is that there are cases in which two incompatible consequents appear to be equally acceptable, so that we lack a criterion to choose one of the two rival conditionals. On the pattern of the famous Bizet-Verdi case proposed by Quine we can build an unlimited number

[1] For a survey of the logic of conditionals from both a historical and theoretical viewpoint see (ARLO-COSTA, 2007).

of examples. For instance, just to remain in the field of Italian opera, one could agree on the truth of the following two statements and on the fact that they transmit an equal amount of information:

(1) Verdi wrote the air “La donna è mobile”

(2) I hate the air “La donna è mobile”

So the following two conditionals appear to be equally acceptable, the first in view of (1) and second in view of (2):

(a) If I had been Verdi I would have written “La donna è mobile”

(b) If I had been Verdi I would not have written “La donna è mobile”

May we qualify as antinomies the paradoxes of the Bizet-Verdi family? An antinomy is by definition a contradiction derived from consistent premises. The notion of derivation which is used here is the one defined in standard first order logic, which offers the framework to reconstruct not only the Liar antinomy but all the well-known antinomies developed during the so-called “Crisis in the Foundations of Mathematics” (Cantor, Burali-Forti, Berry, Russell ...). But the implication relation occurring in (a) and (b) and usually represented by the conditional operator symbolized by “>” does not refer to the logical relation of derivation as above defined. The corner operator is not even reducible to strict implication, i.e. to necessary implication. David Lewis characterized counterfactual conditionals as *variably strict* conditionals, this meaning that in $A > B$ the consequent B follows not from A but from A conjoined with the “quasi-variable” $w(A)$, where $w(A)$ represents the conjunction of the most informative elements of the Background Knowledge compatible with A. If “$\supset$”stands for the material conditional and “$\Box$” for the necessity operator as axiomatized in KT, then a definition of the corner operator might be given as follows:

(Def >) $A > B =_{df} \Box\, ((w(A) \wedge A) \supset B)$

The minimal requirement we may put on $w(A)$is provided by the axiom

$(\Diamond\, w)\ \Diamond A \supset \Diamond\, (w(A) \wedge A)$

So the paradoxes of the Bizet-Verdi family are not properly speaking antinomies. The same could be said of other counterfactual paradoxes known in the literature. For instance, the following one, which stems from a puzzling question which was loved by Abraham Lincoln:

(3) If the tail of a dog were called "paw", how many paws would have a dog?

The antecedent of counterfactual question conjectures the existence of a language which is different form the one employed in the very formulation of the question itself. This class of counterfactuals could be called class of *Counterlinguistic Conditionals*.
Two incompatible answers to the question (3) seem to be equally plausible:

(4a) If the tail of a dog were called "paw", the dog would have five paws

(4b) Even if the tail of a dog were called "paw", the dog would have four paws.

Other paradoxes of the counterfactuals stem from a supposed illegitimacy of the antecedent. This problem is especially serious for consequentialists inasmuch the derivation of the consequent, from their viewpoint, is essentially granted by a certain set of natural laws L_1, L_2.... Let us suppose that the conjunction of laws is a certain theory CL. What happens if we suppose just the falsity of CL or of some of its components? We are in front of the conditionals which Goodman calls –ɜ conditionals. The answers given to the puzzle have been various. In the first place, we are allowed to draw conclusions from CL provided that in the derivation no use is made of laws depending on CL. But if we need such laws to draw a counterfactual conclusion we have to do with a self-defeating argument. In this case one may choose, for instance, to consider all counterlegals false or to deny that they have a truth-value[2].

The problem of counterlegal antecedents has actually some kinship with the Liar antinomy, but is not stricly speaking an antinomy. We may even make the absurd supposition that all the known laws of nature are false, but on this premise one could anyway draw consistent

[2] The former solution is implicit in Goodman's theory, while the latter has been proposed by Reichenbach

conclusions by using simply the logical laws of the background logical system. Making a further step, supposing the falsity of some law of logic L_i (e.g. the Excluded Middle) means to hypothesize a contradiction only if the background logical system S contains L_i , but not in any weaker system S' not including L_i.

§2. The question we now want to discuss is whether a Liar-like antinomy exists which can be written in a conditional language. As is well known, the Liar should not be confused with the Epimenides Paradox or its variants which belong to the class of what is called the family of Prior's Paradoxes[3]. Furthermore, the Liar has many variants, some of which may be introduced as generalizations of it. An example is the so-called "Card Paradox", originally due to the English logician Philip Jourdain, which can be simplified as follows:

$\mathbf{a}_1$: The statement written in the next line is true

$\mathbf{a}_2$: The statement written in the preceding line is false It is pointless to consider here that the two-vertices graph containing a 1 and a 2 may be generalized to a closed figure having an even number of vertices, connected by arrows, in which there is an alternation of statements of the following kind:

$\mathbf{a}_1$, $\mathbf{a}_3$, $\mathbf{a}_5$ ... The statement written at the next vertex is true

$\mathbf{a}_2$, $\mathbf{a}_4$, $\mathbf{a}_6$ ... The statement written at the next vertex is false

It is easy to show, for every closed figure of a finite length having the given pattern, that the statements in the graph yield an antinomy[4]. In case of a figure in which the vertices a_n and a_{n+1} are identical, i.e. $a_n = a_{n+1}$, there is not one but two possible self-referential statements: the standard Liar ("this statement is false") and the standard Truth teller ("this statements is true").

Another variant of the Liar is "Pinocchio's Paradox", which relies on the acquired knowledge of the fact that Pinocchio's nose grows if and only if he is lying. The paradox is provided

[3] See (PRIOR, 1961).

[4] It is to be noted that an antinomy follows even if the predicates "true"' and "false"' in the definition of the figure are inverted. An antinomy also follows from simply stipulating that the truth value of the even vertices is different from the truth value of the odd vertices.

by the statement

(5) Pinocchio says: “my nose is growing now”.

Is Pinocchio lying or not[5]?

Let us begin by trying to define a conditional version of the Card Paradox.

The first step is to imagine that the two sides of the card contain conditional statements, not categorical statements. In order to simplify the exposition we will assume that the background system is a minimal system of classical conditional logic. We choose the system named by Chellas CK [6] extended with ID +MP, i.e. the two axioms

(ID): $A > A$

(CMP): $(A > B) \supset (A \supset B)$

We will also assume a definition of the necessity operator in conditional terms

(Def $\Box$)$\Box A =_{df} \neg A > A$

which is of course equivalent to $\Box A =_{df} \neg A > \bot$

As a consequence of the Conditional Modus Ponens (CMP), the behaviour of $\Box$ includes the law T: $\Box A \supset A$. We will then also assume that the logic of $\Box$ is provided by the normal system KT.

Let us now define two statements X and Y in the following way:

[5] See (ELDRIDGE-SMITH, 2010).

[6] The axioms of CK, subjoined to the standard PC-calculus, are $(A > (B \wedge B')) \supset (A > B \wedge A > B')$, $(A > B \wedge A > B') \supset A > (B \wedge B')), A > \bot$. The rules are $A \equiv A'/A > B \supset A' > B$ and $B_1 \wedge B_2 \wedge ...B_n/A > B_1,...,A > B_n$. See (CHELLAS, 1975).

X: X is true > Y is true

Y: Y is true > X is false

It is inessential (but we can do it) to suppose the existence of a couple of points a_1 and a_2 (the card) such that a_1 is associated to X and a_2 is associated to Y. > is not a transitive relation, so we are not allowed to derive by transitivity the conclusion that X is true > X is false. Due to CMP, however, an antinomy is easily derived. In fact from X and Y we have straightforwardly

X': X is true ⊃ Y is true

Y': Y is true ⊃ X is false.

Now if X is true, by CMP X' is true. But if X' is true, given that X is by hypothesis true, Y is true by standard Modus Ponens. Thus Y' is also true, so by the antecedent of Y' and standard Modus Ponens it follows that X is false: contradiction. Since the supposition that X is true leads to a contradiction, this amounts to a proof of $\neg X$.

(b) Suppose X is false. Here we face the problem of establishing what we intend by negating a conditional. In order to avoid the introduction of a possible world semantics, we may simply look at the definition of > in terms of quasi-variables proposed at p. 1. According to such a definition the negation of X, i.e. $\neg(X$ is true $> Y$ is true), amounts to $\Diamond((w(X) \wedge$ X) true and Y false). So we have

(6) $\neg X \vdash \Diamond(w(X) \wedge X \wedge \neg Y)$

and by KT

(7) $\Diamond(w(X) \wedge X \wedge \neg Y) \vdash \Diamond X$

So by (6) and (7)

(8) $\neg X \vdash \Diamond X$

Now we showed before that X is provably false, so $\neg X$ is provably true. So by Necessitation $\Box\neg X$ is also provable. Then by (8) and Adjunction $\neg X \vdash \Diamond X \wedge \Box\neg X$: contradiction.

If we pass from the Card Paradox to the standard Liar we have to do with the statement which says of itself that, if it is true, it is false:

X: X is true > X is false

The conclusion to be drawn here is especially simple since $X > \neg X$ is by definition equivalent to $\Box \neg X$. X is then equivalent to stating that X is necessarily false. So if X is true, X is necessarily false, which yields a contradiction by the instance of T $\Box\neg X \supset \neg X$. So there is a proof that X is false, i.e. a proof of $\neg X$. If there is a proof that X is false then $\Box\neg X$ is a theorem by Necessitation. But $\Box\neg X$ is $X > \neg X$, so X: contradiction.[7]

A variant of X which should be taken into account is what one obtains in exchanging true and false in X:

X-: X- is false > X- is true

According to the definition of $\Box$, X- amounts to saying of itself that is necessarily true, which obviously does not lead to any contradiction. We could consider this assertion a conditional variant of the Truth Teller. Suppose in fact that it is false: then $\Diamond(w(\neg X) \wedge \neg X)$ and $\Diamond\neg X$, which is not only consistent with X- but in KT logically follows from the supposition $\neg X$.

§3. The systems of classical conditional logic axiomatize an operator > which has an intermediate strength between the material conditional and the strict conditional -3. As already remarked, we will have among the theses the wff $\perp > A$. But according to the consequentialist viewpoint which has been outlined at the beginning, this formula cannot be logically valid. From a consequentialist viewpoint, in fact, one should accept what is sometimes called Strawson's thesis or (weak) Boethius' Thesis:

(BT) $A > B \supset \neg(A > \neg B)$

(BT) states that $A > B$ e $A > \neg B$) cannot be both true. This fact is conflicting with the thesis $\perp > A$, since $\perp > A$ and $\perp > \neg A$ are both valid.

From (BT), via the identity law $A > A$, one obtains by Uniform Substitution and Modus Ponens the so-called Aristotle's Thesis

[7] The proof has the same structure of the one provided for the so-called Knower Paradox in (PRIEST, 1991, p.196)

(AT) $\neg(A > \neg A)$.

In any system with Uniform Substitution and the law $\neg\neg AA$, (AT) is equivalent to the following:

(AT-) $\neg(\neg A > A)$

From both AT and AT- it follows that $A > \neg A$ and $\neg A > A$ are logical contradictions and that they are both equivalent to $\perp$ so provably PC- equivalent.

If > has such properties, the conditional Liar turns out to be especially simple. In this new interpretation of the corner operator the statement

X*: X* is true > X* is false

turns out to be equivalent to

X*- : X* is false > X* is true

So the distinction between the Liar in the two forms (X) and (X-)(see page 6) here disappears. What it turns out is that both X* and X*- are simply contradictions, being both equivalent to $\perp$.

We should open a reflection on the problem whether that **X*** or **X*-** are actually antinomies. As a matter of fact, they have the logical form of a wff which is the negation of a thesis w.r.t. the background system, i.e. of a contradiction. This cannot be said of $X : X > \neg X$. If the background system is CK, X is equivalent to $\Box\neg X$, which lacks the logical form of a contradiction.

§4. An objection to the treatment of the conditional antinomies outlined in §3 is that it can be reproduced for any conditional operator satisfying the minimal conditions required for >. It is easy to show, for instance, that the axioms of CK+ID+MP are theorems if the conditional operator is replaced by the operator of strict implication. A further step consists in recalling that the box operator is defined in systems of strict implication as

(Def $\Box$) $\Box A =_{df} \neg A \prec A$

which allows concluding that $\neg A \prec A$ in classical conditional systems turns out to be equivalent to $\neg A > A$. So the proposed conditional Liar (X) is equivalent to the same statement in which $>$ is replaced by $\prec$.

A point of discrimination between $>$ and $\prec$ would be made by introducing for $>$ such strong axioms as $(A \wedge B) \supset \neg A > B$ or the so-called Conditional Excluded Middle $A > B \vee A > \neg B$: both principles do not hold for strict implication, but they are irrelevant to build the above reported antinomic argument.

We recall that the conceptual distinction between conditional statements and statements of strict implication is that the truth of the former (especially when they are counterfactual) may depend on a set of presuppositions compatible with the antecedent, which, as we saw, is here formally represented by the quasi-variables **w**(A), **w**(B), **w**(C) ... conjoined to the antecedents of the conditional. Due to the occurrence of such contextual elements of information, the conditional operator $>$ lacks some of the properties of strict implication, and especially Monotonicity, Contraposition and Transitivity.

A problem which deserves attention now is to find an antinomic argument which makes use of the $>$ - operator which is unequivocally endowed with the properties of a context-dependent operator.

To begin with, we could see whether there are ways to convert categorical statements into statements with the same meaning involving context-dependent conditionals. As a matter of fact, there is a variety of ways to perform such a translation. Let us take this example:

(9) My cat is black

We list here four ways to paraphrase this statement into one involving context- dependent conditionals:

i) Anyone who were to see my cat would see that it is black

ii) Any object that were identical to my cat is black

iii) If there is no error in what I am saying, my cat is back

iv) I would be a liar if I said that my cat is not black.

On this set of proposed translation the remarks are as follows:

a) The paraphrase in i) is inspired to the so-called phenomenalism as proposed originally by Stuart Mill. It applies only to statements which are reports of observation and not to other kinds of statements, e.g. to logical or analytical propositions. However, it could be modified in various ways, e.g. by saying "Anyone who were to know all the properties of my cat would know that it is black". But only an omniscient ideal subject may know *all* the properties of my cat, so that the antecedent supposes an impossibility and may be qualified as a particular counterlegal supposition (see p. 2).

b) The conditionals occurring in paraphrase of class 2) are called by Goodman *counteridenticals*. The paraphrase could be applied also to non-observational statements by use of propositional quantification. E.g. $\forall p$ (if p were a proposition identical to 2+2=4 it would be true). However, here the operator > has the same sense of necessary implication, so we are back to the trivialization deplored in the preceding sections.

c) The statement 3) has not exactly the sense of "my cat is black". Its sense is that the truth of what is asserted depends on the supposition that there are no mistakes in what I am saying. Its sense is more plausibly rendered by "it is likely that my cat is black".

d) The paraphrase 4) appears to be the more proper rendering of (9). To argue for this conclusion, we simply reason as follows. If it is true that my cat is black, then I would lie if I were to deny this fact, so the statement in 4) is true. If it is false, I would say something true in asserting that my cat is black, so I would say something false in denying it.

The formal rendering of the paraphrase in 4) is not simple. We will follow Prior (1961) in introducing a variable d for monadic functors appliable to propositions, of the kind which he uses in his reconstruction of the Epimenides argument. Here we will use the same symbol, with the meaning of "a says that -"[8] . This operator is simply governed by the axiom

(Axd) $dA \supset \neg d \neg A$

[8] The term "a" stand for any proper name of a human being, including the speaker (so "I" could also replace a).

We can also introduce two constant **F** and **T** with the meaning of False and True, such that $d\mathbf{F}$ means "a says a Falsity", "a says the Truth". A mixed axiom involving d, **F** and **T** is the following:

(LEM) $\neg(dA > d\mathbf{F}) \equiv d(A) > d\mathbf{T}$

We may write L (for "*a* is lying" or "*a* is a liar") in place of $d\mathbf{F}$ and $\neg\mathbf{L}$ in place of $d\mathbf{T}$.

Let us now introduce an axiom which expresses the translation of any statement A in a context-depending conditional having the same meaning of A:

(Ax $\equiv$) $A \equiv d(\neg A) > \mathbf{L}$

By uniform substitution in Ax$\equiv$ then:

(Ax$\equiv$') $\neg A \equiv dA > \mathbf{L}$

A noteworthy property of (Ax$\equiv$) is that it would be untenable if $>$ were replaced by $⥽$. In fact in this case we could perform the following proof:

(10) $A \supset (d\neg A ⥽ \mathbf{L})$

then by Monotonicity of $⥽$

(11) $A \supset ((d\neg A \wedge \neg A) ⥽ \mathbf{L})$

But if A is true (11) is a false statement. In fact $(d\neg A \wedge \neg A) ⥽ \mathbf{L}$ means that if A is false and I am saying that A is false this means that I am lying. This is clearly wrong since $d\neg A \wedge \neg A$ implies that I tell the truth by saying $\neg A$. So A implies a falsity. The relevant point is that the step from (10) to (11) cannot be done with $>$ in place of $⥽$, since $>$ is a non-monotonic operator and does not allow so-called Weakening[9]

[9] In some systems of classical conditional logic Weakening is allowed with some restriction, for instance in the weak form $\neg(A > \neg B) \supset (A > C \supset ((A \wedge B) > C)$. This theorem is however not at disposal in CK + Id +MP.

Let us now examine the self-referential statement X whose intuitive meaning is "If a were to say the present conditional a would lie"

(12) $X : dX >$**L**

Now substituting X to A in the equivalence Ax≡' one reaches the following equivalence

(13) $\neg X : dX >$**L**

But according to (Ax ≡) $dX >$ **L** is X, so by (13) and replacement we reach the contradiction

(14) $X \equiv \neg X$

The contradiction may then constructed as the conclusion of a formal argument.

However, we may also reach the same result via an intuitive argument along the lines of the standard antinomic reasoning. The argument relies on a tacit premise: that supposing A is the same as supposing that someone (e.g. I myself) says A, which is indeed plausible but depends on the meaning of the terms "supposing" and "saying".

(a) Let us suppose that X, i.e. $dX >$ **L** is true. Thus we are supposing that I am saying $X(dX)$, so by Conditional Modus Ponens **L**, i.e. that I am saying something false. So by supposing that X is true I conclude that I am a liar in saying this. Contradiction.

(b) Let us suppose that X, i.e. $dX >$ **L**, is false. This implies implies supposing the truth of $dX > d$**T** by axiom LEM. Then I would tell the truth in saying that X is true. Contradiction.

§5. To conclude this note, we may show that it is possible to build in conditional terms two well-known variants of the Liar which are the Epimenides and the Truth Teller.

Epimenides the Cretan says

(15) All Cretans are liars.

In the conditional paraphrase we have proposed above (15) means:

(16) If Epimenides said that some Cretan tell the truth, Epimenides would be lying

Suppose Epimenides tells the truth. Then there is some Cretan (i.e. Epimenides) who tells the truth. So if Epimenids makes the assertion that some Cretan tells the truth he is telling the truth.Then by Conditional Modus Ponens applied to (16) it follows that he is lying. Contradiction.

Suppose Epimenides tells a falsehood. Then it is possible that, *ceteris paribus*, Epimenides says that some Cretan tell the truth and that Epimenides tells the truth in making this assertion. This is not a contradiction unless there is only one Cretan in the world, i.e. Epimenides.

So, given that the first alternative implies a contradiction and that Epimenides is not the only Cretan, Epimenides makes a false assertion.

The so-called Truth Teller is the statement:

(17) What I am saying is true or equivalently

(18) Xt: Xt is a true statement.

In conditional terms the rendering of (17) is

(19) I would be lying if I said that the present statement is false and the rendering of (18) is

(20) I would be lying if I said that Xt> is false

Suppose Xt> is true. Then the conditional (20) is also true. Suppose Xt> it is false: this means that it logically possible that Xt> is false and that I am telling the truth about this fact, which is the negation of (20).

Then what we may say about (20) is that it is true if and only if it is true, and that there is no procedure to establish if is true or false. (20) is then an undecidable statement, just as the standard Truth Teller.

References

ARLO-COSTA, H. The logic of conditionals. *Stanford Enc. Of Philosophy*, 2007. Disponível em: <http://stanford.library.usyd.edu.au/archives/sum2010/entries/logic-conditionals/>.

CHELLAS, B. Basic conditional logic. *Journal of Phil. Log.*, v. 4, p. 133–154, 1975.

ELDRIDGE-SMITH, P. The pinocchio paradox. *Analysis*, v. 70, p. 212–215, 2010.

MCCALL, S. Connexive logic. *Stanford Enc. Of Philosophy*, 2007. Disponível em: <http://http://plato.stanford.edu/entries/logic-connexive/>.

PRIEST, G. Intensional paradoxes. *Notre Dame Journal of Formal Logic*, v. 32, p. 193–211, 1991.

PRIOR, A. N. On a family of paradoxes. *Notre Dame Journal of Formal Logic*, v. 2, p. 16–32, 1961.

TERMOS SINGULARES INDEFINIDOS: FREGE, RUSSELL E A TRADIÇÃO MATEMÁTICA

Daniel Durante Pereira Alves *

Resumo: É bem conhecida a divergência entre as posições de Gottlob Frege e Bertrand Russell com relação ao tratamento semântico dado a sentenças contendo termos singulares indefinidos, ou seja, termos singulares sem referência ou com referência ambígua, tais como '*Papai Noel*' ou '*o atual rei da França*' ou '$\frac{1}{0}$' ou '$\sqrt{4}$' ou '*o autor de Principia Mathematica*'. Para Frege, as sentenças da linguagem natural que contêm termos indefinidos não formam declarações e portanto não são nem verdadeiras nem falsas. Já para as sentenças da matemática, Frege defende que elas precisam ser corrigidas através da convenção forçada de uma referência não ambígua. Russell, por outro lado, aceita os termos indefinidos e propõe, através de sua teoria das descrições definidas, uma maneira de avaliar as sentenças em que eles ocorrem; e Quine amplia a teoria de Russell para abranger também os nomes com problemas de referência. Na prática da matemática são comuns os termos singulares indefinidos, sem referência, tais como '$\frac{1}{0}$', ou com referência ambígua, tais como '$\sqrt{4}$'. Apesar de não haver uma sistematização rigorosa desta situação entre os matemáticos, há, no entanto, um conjunto de regras convencionais que tradicionalmente costumam ser aplicadas no tratamento matemático dos termos indefinidos. Nossa proposta é tomar a convenção matemática como inspiração e modelo para apresentar uma interpretação semântica formal para as descrições definidas e os nomes e utilizá-la como um argumento que favorece a abordagem de Russell relativamente à de Frege.[1]

Palavras-chave: Semântica. Descrições definidas. Referência. Funções parciais. Nomes. Termos singulares.

Abstract: It is well known the difference between the positions Gottlob Frege and Bertrand Russell regarding the semantic treatment of sentences containing indefinite singular terms, which are terms without or with ambiguous references, such as '*Santa Claus*' or '*the present King of France*' or '$\frac{1}{0}$', or '$\sqrt{4}$', or '*the author of Principia Mathematica*'. For Frege, sentences of natural language containing undefined terms do not form statements and therefore they are neither true nor false. As for the sentences of mathematics, Frege argues that they need to be corrected by a forced convention settling to them an unambiguous reference. Russell, on the other hand, accepts undefined terms and proposes a way to evaluate sentences in which they occur through his theory of definite descriptions; and Quine extends Russell's theory to encompass also the names with problems of reference. In mathematical practice, singular terms without reference, such as '$\frac{1}{0}$', or with ambiguous references, such as '$\sqrt{4}$', are very common. Although there is no rigorous systematization to deal with them among mathematicians, there is, however, a conventional set of rules traditionally applied in the mathematical treatment of undefined terms. Our proposal is to take the mathematical convention as an inspiration and model to present a formal semantic interpretation for definite descriptions and names and use it as an argument favoring Russell's over Frege's approach.

Keywords: Semantic. Definite descriptions. Reference. Partial funcions. Names. Singular terms.

*Professor de Filosofia da Universidade Federal do Rio Grande do Norte. E-mail: *durante@ufrnet.br.*

[1] Gostaria de agradecer ao Prof. José Eduardo Moura, pela cuidadosa leitura, correções e sugestões.

Introdução

Não é incomum encontrarmos nos discursos *termos singulares indefinidos*, ou seja, expressões com a função sintática de termos singulares, mas que ou não se referem a nada ou referem-se ambiguamente a mais de um objeto ou indivíduo. Não temos qualquer problema, por exemplo, em entender o significado da expressão '*o atual rei da França*'. Sabemos que ela é uma *descrição definida*, uma descrição que, por iniciar-se com um artigo definido, espera-se que denote uma única e específica pessoa. O problema é que a França é uma república e, portanto, não tem rei. Então não há nada que seja a referência da expressão '*o atual rei da França*'.

Outros exemplos de termos singulares indefinidos são:

- *O autor de 'Principia Mathematica'* [2]
- *O nono planeta do sistema solar* [3]
- *Papai Noel* [4]
- $\frac{1}{0}$ [5]
- $\sqrt{4}$ [6]

Há algumas questões interessantes que os termos singualres indefinidos suscitam. Considere as seguintes sentenças:

- *O atual rei da França é careca.*
- *O autor de 'Principia Mathematica' ganhou um prêmio Nobel.*
- *O nono planeta do sistema solar chama-se Plutão.*
- *Papai Noel mora no polo norte.*
- $\frac{1}{0} = \frac{1}{0}$
- $\sqrt{4} > 0$

Como todas estas sentenças possuem termos singulares indefinidos, elas nos impõem algumas questões: estamos, de fato, dizendo alguma coisa quando as proferimos ou elas são apenas erros linguísticos? Se elas são erros linguísticos, são erros de que natureza? Não são erros sintáticos, já que são expressões gramaticalmente bem formadas. Seriam então erros semânticos? Mas o que é, exatamente, um erro semântico? É erro de quem profere ou erro de quem interpreta a sentença? Neste caso, o que dizer da terceira sentença, que não seria erro linguístico

[2] Como a obra *Principia Mathematica* tem dois autores, Russell e Whitehead, então a descrição é um termo singular indefinido pois tem forma sintática de termo singular, mas denota ambiguamente mais de um indivíduo.

[3] Lembrando que havia 9 planetas no sistema solar, mas há alguns anos Plutão deixou de ser considerado planeta e, portanto, hoje só há 8.

[4] Desconsiderando-se aqui qualquer sentido especial ou cultural de existência.

[5] Lembrando que não há divisão por zero.

[6] Como tanto $2 \times 2 = 4$ quanto $-2 \times -2 = 4$, então $\sqrt{4}$ denota ambiguamente tanto 2 quanto -2.

até há poucos anos, mas hoje seria. Por outro lado, se estas sentenças não são erros linguísticos, como todas elas são declarações, sentenças que afirmam algo, então sobre todas elas cabe julgarmos se são verdadeiras ou falsas. Mas como fazemos isso? Se, por exemplo, consideramos falsa a primeira sentença – '*O atual rei da França é careca*', então, logicamente, a sua suposta negação, a sentença '*O atual rei da França não é careca*', deve ser verdadeira. Mas qual o fundamento para considerarmos verdadeira ou falsa qualquer destas sentenças, já que não há nenhum atual rei da França ou nono planeta do sistema solar? Se sentenças declarativas com termos indefinidos não são erros linguísticos então, sendo declarações, é preciso que haja algum critério de verdade que se aplique a elas. Como definir este critério? Qual seria o seu fundamento?

Além disso, as duas últimas sentenças da lista acima atestam que este não é um problema apenas da linguagem ordinária, mas ocorre também na matemática. Elas são verdadeiras, falsas ou simplesmente estão matematicamente mal escritas, e por isso não se configuram como proposições? Apesar de não haver um tratamento matemático sistematizado para questões deste tipo, há, sim, conforme veremos mais adiante, um conjunto de regras convencionais tradicionais, amplamente aceitas, que costumam ser empregadas para lidar com termos indefinidos na matemática.

O problema de sentenças contendo termos singulares indefinidos já é bastante conhecido e estudado pela tradição analítica da filosofia. São famosas as abordagens divergentes de Gottlob Frege e Bertrand Russell a esta questão. Nosso objetivo neste artigo é apresentar um tratamento linguístico formal para os termos singulares indefinidos de tal modo que seja possível aplicar as regras da convenção matemática também a expressões da linguagem comum. Ao fazer isso, veremos que o tratamento matemático da questão coincide com a abordagem de Russell ampliada por considerações de Quine. Utilizaremos este fato como fundamento de um argumento em favor da abordagem de Russell relativamente à de Frege, ao mostrarmos que a tradição matemática dá suporte a Russell. Para além desta contribuição para o debate histórico, acreditamos que a reformulação em termos funcionais das descrições definidas e nomes próprios que apresentaremos aqui ajuda a esclarecer os princípios que fundamentam a posição Russell–Quine e a mostrar que são estes mesmos princípios que também fundamentam a posição da tradição matemática.

Nossos passos serão, então, os seguintes: em primeiro lugar apresentaremos de modo bastante resumido as abordagens de Frege e Russell do problema linguístico dos termos singulares indefinidos e ampliaremos a abordagem de Russell com as considerações de Quine. Em seguida, apresentaremos o modo como a tradição matemática lida com o problema. Depois proporemos uma nova interpretação formal para nomes próprios e descrições definidas como termos funcionais. Então, veremos que as nossas redefinições funcionais nos habilitam a aplicarmos as regras convencionais da tradição matemática a sentenças da linguagem comum e que, ao fazermos isso, obtemos os mesmos resultados que a abordagem Russell–Quine obtém. Por fim, explicitaremos os fundamentos que regem tanto a abordagem Russell–Quine quanto a tradição matemática, concluindo nosso argumento. Como subproduto das considerações feitas, terminamos o artigo com um breve comentário que evidencia um aspecto de inadequação da lógica clássica para lidar com os problemas que os termos indefinidos impõem à matemática.

A Abordagem de Frege

Em seu seminal artigo de 1892, *Sobre o Sentido e a Referência*, Frege apresenta a famosa distinção entre *sentido* e *referência* das expressões linguísticas:

> É, pois, plausível pensar que exista, unido a um sinal (nome, combinação de palavras, letras), além daquilo por ele designado, que pode ser chamado de sua referência (*Bedeutung*), ainda o que eu gostaria de chamar de o sentido (*Sinn*) do sinal, onde está contido o modo de apresentação do objeto. (FREGE, 2009, 131)

Assim, as expressões '*Aristóteles*', '*o Estagirita*' e '*o fundador do Liceu*', por exemplo, têm todas a mesma referência, Aristóteles, mas têm sentidos diferentes, pois são modos de apresentação diferentes do mesmo objeto/indivíduo.

Tanto partes de uma sentença têm sentido e referência, quanto a própria sentença completa. Os termos individuais, ou nomes próprios, terão como referência o objeto nomeado e como sentido o modo de apresentação do objeto. Mas qual o sentido e a referência de uma sentença completa? Frege chamou de *pensamento* (*Gedanke*) o sentido de uma sentença completa (FREGE, 2009, 137) e de *valor de verdade* a sua referência (FREGE, 2009, 139). E entre os valores de verdade distinguia apenas duas possibilidades: *o verdadeiro* e *o falso*. Dessa forma, quaisquer sentenças verdadeiras, tais como '*Aristóteles foi discípulo de Platão*' e '*A água ferve a cem graus célsius ao nível do mar*' têm a mesma referência: o verdadeiro. Elas são distintas apenas em seus sentidos, ou seja, nos pensamentos que exprimem.

Uma discussão mais aprofundada sobre a estranheza em considerar valores de verdade como referência de sentenças ou sobre o que vem a ser exatamente o sentido (modo de apresentação e pensamento) e qual o seu fundamento ontológico, está além dos nossos interesses específicos. Para os nossos propósitos é suficiente notarmos que a referência de uma expressão é aquilo (a coisa) de que se está falando quando se usa tal expressão, e que por isso, expressões como '*o atual rei da França*' não têm referência. Já o sentido de uma expressão liga-se ao que normalmente chamamos de significado. É aquilo que falantes e ouvintes entendem pela expressão e que os habilita a identificar sua referência (MENDELSOHN, 2005, XV). Ou ainda, o sentido "é o que é comunicado ou carregado pela expressão, a informação que ela contém." (MENDELSOHN, 2005, 35) [7]

Para Frege, tanto o sentido quanto a referência das sentenças completas são obtidos por composição dos sentidos e referências de suas partes. Tal constatação passou a ser conhecida como princípio da *composicionalidade* do sentido e da referência (MENDELSOHN, 2005, 11 e 37). Frege não apresenta explicitamente este princípio em seu texto, mas ele pode ser inferido de seus escritos principalmente pelo uso de um outro princípio que pressupõe logicamente a composicionalidade. É o princípio que Carnap (1947, 122) chamou de *substitutibilidade*, que afirma que a substituição de partes de uma sentença por outras que tenham a mesma referência (ou sentido) não afeta a referência (ou sentido) da sentença completa. Os dois trechos abaixo são exemplos que atestam a adesão de Frege aos princípios da substitutibilidade e composicionalidade tanto para o sentido quanto para a referência das sentenças.

> Se nossa suposição é correta, de que a referência de uma sentença é seu valor de verdade, então este tem de permanecer inalterado, se uma parte da sentença for substituída por

[7] A teoria de Frege sobre o sentido e a referência, base de seu tratamento aos termos indefinidos, além de ter muitos detalhes, suscitou toda uma literatura secundária com críticas, complementos e desenvolvimentos, dentre os quais, alguns dos mais influentes são: (RUSSELL, 1905), (CHURCH, 1996), (CARNAP, 1947), (KAPLAN, 1979), (DUMMETT, 1993), (MENDELSOHN, 2005).

uma expressão que tenha a mesma referência, ainda que sentido diverso. (FREGE, 2009, 140)

> Se tudo quanto importa fosse apenas o sentido da sentença, fosse apenas o pensamento, então seria desnecessário preocupar-se com a referência de uma parte da sentença; pois para o sentido da sentença somente importa o sentido desta parte. (FREGE, 2009, 138)

Mas se a referência de uma sentença completa, seu valor de verdade, é obtida por uma composição das referências de suas partes, o que ocorre com a referência da sentença quando alguma de suas partes for um termo indefinido sem referência ou com referência ambígua? Frege responde a esta pergunta afirmando que a falta ou ambiguidade da referência de algum de seus componentes faz com que a sentença resultante, ela própria, não tenha referência (FREGE, 2009, 147). Então, em princípio, para Frege, as sentenças com termos singulares indefinidos que vimos acima não seriam nem verdadeiras nem falsas.

Chamaremos a esta concepção de Frege de que as sentenças contendo termos singulares indefinidos não são nem verdadeiras nem falsas de *Teoria Frege–Strawson*. Frege trata esta questão como uma falha das linguagens ordinárias e defende que em uma linguagem logicamente perfeita (a sua conceitografia) não deve ser possível nem introduzir nomes próprios que não tenham referência, nem que qualquer expressão gramaticalmente correta "construída como um nome próprio" careça de referência ou tenha referência ambígua (FREGE, 2009, 147). Estes preceitos de Frege vingaram na lógica de predicados clássica, onde, para garantir que todos os termos individuais tenham referência, exige-se, em todos os modelos, que as constantes individuais (os nomes) tenham referência definida e que os símbolos de função sejam interpretados por funções totais. Como consequência disso, não existem termos indefinidos na lógica clássica e qualquer modelo atribui valor de verdade a todas as sentenças fechadas.

Mas ainda que proibidos na lógica clássica, os termos indefinidos ocorrem na matemática, e Frege era tanto consciente deste fato, quanto acreditava na necessidade de dar-lhe um tratamento, de corrigir esta "falha". Ele propõe duas maneiras alternativas de corrigir este problema no contexto da matemática. Uma que ele sugere no próprio artigo *Sobre o Sentido e a Referência*, que chamaremos *Teoria Frege–Carnap*, e a outra que ele aponta no §11 das suas *Leis Básicas da Aritmética*, que será denominada de *Teoria Frege–Grundgesetze*. Estes três nomes para as divergências na abordagem de Frege: *teoria Frege–Strawson*, *teoria Frege–Carnap* e *teoria Frege–Grundgesetze* foram propostos por (KAPLAN, 1972) e adotados por (PELLETIER; LINSKY, 2009).

A correção que Frege propõe em *Sobre o Sentido e a Referência*, a teoria Frege–Carnap, consiste em convencionar uma atribuição artificial de referência a qualquer termo indefinido que ocorra no contexto da matemática:

> Considero igualmente oportuno se precaver contra os nomes próprios aparentes carentes de toda referência. A história da matemática narra erros que se originaram dessa maneira (FREGE, 2009, 147). [...] De acordo com as observações acima, uma tal expressão deve sempre ter assegurada uma referência por meio de uma convenção especial, por exemplo de que sua referência será o número 0 se nenhum objeto, ou mais de um, cai sob o conceito (FREGE, 2009, 147 – nota 58).

A correção proposta nas *Leis Básicas da Aritmética*, teoria Frege–Grundgesetze, evidentemente restrita ao contexto da matemática, consiste, por sua vez, em uma convenção alternativa, um pouco mais elaborada, de atribuição artificial de referência. No §11 Frege introduz uma função que ele denomina de *substituto para o artigo definido* que, basicamente, atua como um desambiguador artificial para expressões do tipo "o x tal que $\varphi(x)$".[8] Quando há um único valor possível de x para a expressão "o x tal que $\varphi(x)$", como por exemplo, "o x tal que $x+3=5$" ou seja, "o número que quando acrescido de 3 é igual a 5", a função *substituto para o artigo definido* retorna este valor de x, que no caso é 2. Quando há mais de um valor possível de x para a expressão "o x tal que $\varphi(x)$", como por exemplo, "o número que multiplicado por si mesmo é igual a 4", ou seja, "o x tal que $x^2 = 4$", a função retorna o conjunto dos valores que satisfazem a expressão, que neste caso é $\{-2, 2\}$. Note que mesmo havendo mais de um valor possível para x neste caso, ao retornar o conjunto destes valores, a função *substituto para o artigo definido* está retornando um único objeto, o conjunto de valores[9] e, portanto, está desambiguando a expressão. Por fim, quando nenhum valor de x satisfaz "o x tal que $\varphi(x)$", como em "o número que multiplicado por 0 é igual a 1", ou seja, "o x tal que $x.0 = 1$", a função retorna o conjunto vazio. (FREGE, 1964, 49–50)

Há muita controvérsia na literatura sobre qual é, de fato, a abordagem que representa a real posição de Frege sobre sentenças com termos indefinidos. Tanto que as referências a Strawson e Carnap nos nomes propostos por Kaplan para as diferentes versões das abordagens de Frege refletem as divergências interpretativas destes autores sobre a posição de Frege que podemos ver em (STRAWSON, 1950) e (CARNAP, 1947).

A estas três teorias poderíamos ainda acrescentar uma quarta, que nem Kaplan nem Pelletier mencionam, mas a qual poderíamos chamar de *teoria Frege–Begriffsschrift*, que corresponde à proposta de reforma da linguagem que ele impôs à conceitografia, e que foi adotada pela lógica clássica de primeira ordem, que simplesmente proíbe a ocorrência de termos indefinidos.

Eu prefiro adotar a interpretação de Strawson, que sustenta que a abordagem principal de Frege sobre o assunto defende que, como consequência do princípio da composicionalidade, sentenças com termos indefinidos não são nem verdadeiras nem falsas. Esta parece ser, pelo menos, a abordagem de Frege com relação à linguagem natural. Há, inclusive, bastante suporte para esta posição em *Sobre o sentido e a referência* (FREGE, 2009, 147). Frege, no entanto, entende esta característica da linguagem natural como um problema, uma falha e suas outras abordagens podem, por isso, ser interpretadas como as suas tentativas de corrigir este problema nos contextos mais restritos da conceitografia e da matemática.

Podemos resumir a abordagem de Frege na seguinte tabela:

Sentença	**Valor de Verdade**
O Atual rei da França é careca	I
Papai Noel não existe.	I
$\frac{1}{0} = \frac{1}{0}$	V

Repare que como as duas primeiras sentenças são da linguagem natural, então aplicamos a abordagem Frege–Strawson, que afirma que sentenças com termos indefinidos têm referência indefinida, ou seja, não têm valor de verdade. Usamos, na tabela, a letra '**I**' de '*valor Indefinido*' para indicar este fato. Por outro lado, a terceira sentença é uma afirmação matemática. Neste

[8] $\varphi(x)$ é uma fórmula sentencial com uma variável livre. Então, a expressão "o x tal que $\varphi(x)$" deveria ter como referência o único indivíduo x que torna a sentença $\varphi(x)$ verdadeira. Acontece que dependendo de qual é a expressão $\varphi(x)$, pode ou não haver x que a torne verdadeira, ou haver mais de um. Nestes casos a expressão "o x tal que $\varphi(x)$" é um termo singular indefinido e precisa ser desambiguada.

[9] *Percurso de valores* na nomenclatura de Frege

caso tanto a abordagem Frege–Carnap quanto a abordagem Frege–Grundgesetze consideram verdadeira a sentença, pois seja qual for a referência atribuída artificialmente por convenção ao termo indefinido $\frac{1}{0}$, esta referência é idêntica a si própria.

Voltaremos outras vezes a esta tabela, conforme a formos complementando com os dados das outras abordagens que apresentaremos. No entanto, já é digno de estranhamento o tratamento de Frege às duas últimas sentenças. A segunda indica que declarações de não-existência supostamente verdadeiras são indefinidas para Frege, e a terceira atribui verdade a uma proposição tradicionalmente considerada falsa.

A Abordagem de Russell

Bertrand Russell, no famoso artigo *On Denoting*, desenvolve a sua *teoria das descrições definidas* com o objetivo de corrigir determinados problemas que ele via nas abordagens até então conhecidas de (MEINONG, 1981) (e seus seguidores como Mally e Ameseder) e (FREGE, 2009). (RUSSELL, 1905, 480)

Russell apresenta a sua abordagem distinguindo a proposição que uma sentença representa de sua expressão verbal. Segundo ele, as descrições definidas não são constituintes legítimas das proposições, mas apenas das expressões verbais em que ocorrem. Deste modo ele propõe a sua teoria como um conjunto de regras de paráfrase que dão a correta interpretação proposicional das sentenças cujas expressões verbais contêm descrições definidas (RUSSELL, 1905, 483). Ele fundamenta esta abordagem afirmando que os problemas com as teorias de Meinong e Frege originam-se, justamente, no tratamento ingênuo que ambos dão às frases denotativas e em especial às descrições definidas ao considerá-las componentes legítimas das proposições, como se fossem nomes.

Não desenvolverei em detalhes as críticas de Russell a Frege e Meinong, nem comentarei sobre a sua pertinência. Estes assuntos têm sido amplamente debatidos em vasta e longeva bibliografia, da qual podemos citar (SEARLE, 1958), (BLACKBURN; CODE, 1978) e (SMITH, 1985). Apenas como uma rápida indicação, podemos dizer que as críticas de Russell a Meinong relacionam-se às consequências lógicas da teoria deste. Segundo Russell, ao considerar qualquer frase denotativa gramaticalmente correta como representativa de um objeto, a teoria de Meinong admite objetos que violam a lei da contradição. Estes objetos são justamente aqueles representados por termos indefinidos sem referência, ou seja, frases denotativas que não denotam, pois '*o atual rei da França*', que não existe, também existiria, segundo Meinong, como referência obrigatória desta frase denotativa. (RUSSELL, 1905, 484)

Quanto a suas críticas a Frege, Russell afirma que a distinção entre *sentido* e *referência*, de Frege, que ele chama de *significado* e *denotação*, evita as contradições acima ao defender que um termo indefinido como '*o atual rei da França*' apesar de ter sentido, não tem referência. No entanto, Russell não aceita o princípio da composicionalidade do sentido defendido por Frege. Ou seja, ele discorda que o significado de uma proposição em que ocorre uma frase denotativa seja uma composição do significado de suas partes. Segundo Russell, quando digo '*O atual rei da França é careca*', não estou predicando sobre o significado de '*o atual rei da França*', mas sobre o indivíduo denotado por esta expressão. Para Russell, se trato a frase denotativa '*o atual rei da França*' como um nome, como o faz Frege, o significado da proposição expressa por '*O atual rei da França é careca*' dependeria da própria denotação da expressão '*o atual rei da França*'. Como tal denotação não existe, a proposição '*O Atual rei da França é careca*' seria sem sentido. Mas ela não é sem sentido (RUSSELL, 1905, 484). É por isso que Russell afirma que não devemos tratar as expressões verbais das descrições definidas como constituintes legítimas de proposições, tal qual o fazemos com os nomes, mas temos sim que parafrasear tais expressões de modo a evitar a armadilha em que Frege caiu.

Outros aspectos da crítica de Russell a Frege dizem respeito à artificialidade do modo como este tentou resolver o problema, convencionando arbitrariamente uma denotação para os termos indefinidos (RUSSELL, 1905, 485) e à dificuldade em defender a diferença entre os conceitos de sentido (significado) e referência (denotação) que, segundo Russell, mediante a necessidade de relacionarem-se um com o outro, acabam por colapsarem em um único e indistinto conceito. (RUSSELL, 1905, 487)

Em termos gerais, a teoria de Russell defende que as descrições definidas devem ser parafraseadas por uma afirmação da *existência* e da *unicidade* do objeto descrito. Assim, quando afirmo que '*O atual rei da França é careca*', estou, para Russell, dizendo que existe um único atual rei da França e que ele é careca. Se simbolizarmos por $\mathsf{Careca}(\mathsf{x})$ e $\mathsf{AtuReiFra}(\mathsf{y})$ os predicados '*ser careca*' e '*ser o atual rei da França*', respectivamente, então uma versão na linguagem da lógica de primeira ordem da interpretação de Russell para a proposição expressa pela sentença '*O atual rei da França é careca*' seria:

$$\exists \mathsf{x}\,(\mathsf{AtuReiFra}(\mathsf{x}) \wedge \forall \mathsf{y}\,(\mathsf{AtuReiFra}(\mathsf{y}) \rightarrow \mathsf{x} = \mathsf{y}) \wedge \mathsf{Careca}(\mathsf{x}))$$

O modo como Russell apresenta esta simples teoria interpretativa é um tanto tortuoso, mas em 1905, antes porém dos *Principia Mathematica*, e antes do estabelecimento de uma convenção de formalização e interpretação para a linguagem da lógica, o modo tortuoso da apresentação de Russell é perfeitamente justificável.

Um primeiro fato a se notar é que, se considerarmos a tradução em linguagem de primeira ordem como uma representação da proposição vinculada à expressão verbal da sentença em linguagem natural, não encontraremos nenhum componente isolável da proposição formalizada que seja a tradução direta de uma descrição definida. Ao invés disso, a presença de uma descrição definida na expressão verbal da sentença indica uma maneira de produzir a proposição, maneira esta que por um lado não exige a existência da entidade supostamente referida pela descrição para que a proposição tenha significado e, por outro lado, envolve a afirmação da existência e da unicidade desta suposta entidade. Este engenhoso expediente habilita Russell a admitir termos indefinidos nas expressões verbais das sentenças sem, no entanto, se comprometer com a existência de entidades abstratas estranhas, que seriam as referências destes termos. Como resultado, não há na proposição resultante de sua interpretação qualquer menção a uma entidade que existe e não existe, como em Meinong, ou que tenha sentido mas não tenha referência, como em Frege. Para Russell, a proposição expressa pela sentença '*O atual rei da França é careca*' simplesmente é falsa. Podemos ver facilmente isto tanto notando que nenhuma substituição de x por qualquer indivíduo tornará verdadeira a fórmula $\mathsf{AtuReiFra}(\mathsf{x})$, ou, de modo ainda mais simples, notando que a paráfrase de Russell é equivalente a afirmação da conjunção entre: (1) '*Existe um único indivíduo que é o atual rei da França*' e (2) '*Este indivíduo, se existir, é careca*'. Como (1) é falsa, a conjunção de (1) com (2) também é falsa.

A interpretação de Russell também se sai bem do embaraço da negação de uma sentença contendo termo indefinido. Tanto a sentença '*O atual rei da França é careca*' quanto sua suposta negação, '*O atual rei da França não é careca*' são consistentemente falsas para Russell. Isto porque a paráfrase de Russell faz com que a segunda proposição não seja a negação da primeira, mas corresponda a conjunção de (1) '*Existe um único indivíduo que é o atual rei da*

França' e (3) '*Este indivíduo, se existir, não é careca*'. Como já vimos, (1) é falsa. Logo, a sentença toda também é falsa. Sua tradução na linguagem de primeira ordem seria:

$$\exists \mathsf{x}\,(\mathsf{AtuReiFra}(\mathsf{x}) \wedge \forall \mathsf{y}\,(\mathsf{AtuReiFra}(\mathsf{y}) \rightarrow \mathsf{x} = \mathsf{y}) \wedge \neg \mathsf{Careca}(\mathsf{x}))$$

O Complemento de Quine à Abordagem de Russell

A teoria de Russell responde bem aos problemas que ele próprio aponta com relação a descrições impróprias (descrições definidas sem referência ou com referência ambígua), mas este mesmo tipo de problema pode ocorrer com nomes impróprios. Qual a interpretação semântica da sentença '*Papai Noel mora no polo norte*'? Não temos aqui uma descrição imprópria, mas um nome impróprio. Um nome para o qual não há indivíduo que seja a sua referência. Meinong e Frege tratam, cada um à sua maneira, esta sentença, mas a teoria da denotação de Russell, por lidar exclusivamente com descrições, nada nos diz sobre nomes impróprios e é, por isso, incompleta.[10] No entanto, se trocarmos o nome '*Papai Noel*' por uma descrição sua, tal como '*o velhinho de barba branca e roupa vermelha que entrega presentes na noite de natal*', podemos aplicar a abordagem de Russell e constatar que a sentença resultante, '*O velhinho de barba branca e roupa vermelha que entrega presentes na noite de natal mora no polo norte*', que é semanticamente equivalente à anterior, é falsa. Seria, portanto, compatível com a teoria de Russell tratar os nomes impróprios de modo assemelhado ao tratamento que ele dá às descrições impróprias. É exatamente isto que (QUINE, 1963) propõe.

Em seu famoso artigo '*Sobre o que Há*', Quine está preocupado em assegurar a possibilidade de se debater sobre ontologia. Para isso dispõe-se a identificar de que forma os discursos se comprometem ontologicamente, ou seja, quando se pode afirmar que uma entidade ontológica específica é subsumida por um determinado discurso. Ele procura se contrapor a um antigo e forte argumento que assume que o uso de um nome torna o discurso comprometido ontologicamente com a suposta entidade nomeada. Este compromisso tornaria inviável qualquer debate ontológico, pois tornaria falsa qualquer afirmação de não existência feita com uso de nomes. Se a interpretação semântica do nome '*Deus*', por exemplo, envolvesse de alguma maneira a entidade nomeada, então a afirmação de que '*Deus não existe*' seria contraditória, pois o uso do nome '*Deus*' nos comprometeria com a existência da entidade que justamente está-se afirmando que não existe.

A teoria de Meinong, segundo Russell, cai exatamente neste tipo de contradição. Já a de Frege, na versão Frege–Strawson que lida com sentenças na linguagem natural, apesar de não se contradizer, não consegue valorar afirmações de não existência supostamente verdadeiras, conforme pudemos notar na tabela do final da seção 2, justamente pela impossibilidade de compor a referência da sentença (seu valor de verdade) quando algum de seus termos não tem referência (o termo referente à entidade que não existe).

Para evitar este problema Quine propõe que os nomes sejam tratados como predicados e que em todas as sentenças em que ocorram, eles sejam substituídos por uma descrição definida envolvendo este predicado. Por fim, ele propõe que estas descrições definidas que substituem os nomes sejam interpretadas semanticamente segundo a teoria de Russell. Assim, o nome '*Papai Noel*', por exemplo, dá origem ao predicado '*ser Papai Noel*' e o seu uso em qualquer sentença

[10] Russell demonstrou, posteriormente, estar ciente deste problema que os nomes impróprios impõem. Tanto que em (RUSSELL, 1985) ele propõe uma radical teoria de eliminação dos nomes.

dá origem à descrição definida '*o indivíduo/objeto que é Papai Noel*'. Então, a sentença '*Papai Noel mora no polo norte*' deve ser parafraseada para '*O indivíduo que é Papai Noel mora no polo norte*'. E esta sentença, por sua vez, deve ser interpretada semanticamente segundo a teoria de Russell, ou seja, deve ser interpretada como a afirmação de que existe um único indivíduo que é Papai Noel e ele mora no polo norte. Como não há nenhum indivíduo/objeto que satisfaça o predicado '*ser Papai Noel*', esta sentença é falsa. Uma tradução sua em linguagem de primeira ordem seria:

$$\exists x\,(\mathsf{EhPapaiNoel}(x) \land \forall y\,(\mathsf{EhPapaiNoel}(y) \to x = y) \land \mathsf{MoraPoloNorte}(x))$$

Afirmações de não existência que envolvem nomes, que eram problemáticas na abordagem de Frege, agora são adequadamente tratadas. A sentença '*Papai Noel não existe*', por exemplo, é parafraseada por '*O indivíduo que é Papai Noel não existe*'. Sua tradução em linguagem de primeira ordem, $\neg\exists x\,\mathsf{EhPapaiNoel}(x)$, é uma sentença verdadeira, já que nenhum x satisfaz o predicado $\mathsf{EhPapaiNoel}(x)$.

Apesar de ter resolvido alguns problemas que a abordagem de Frege enfrentava, há na literatura muitas críticas a esta interpretação descritivista dos nomes. Como anteriormente, não vou detalhar a questão, presente em vasta bibliografia, da qual podemos citar (ALSTON, 1958), (LEWIS, 1973), (STRAWSON, 1961), (KRIPKE, 1980), (CHATEAUBRIAND, 2003). Apenas de modo muito breve indico que a abordagem de Quine se baseia em um tratamento semântico descritivista dos nomes. Segundo esta abordagem, o que nos habilita a entender o uso de um nome, ou seja, a identificar a sua referência, são as descrições que temos do objeto nomeado. Assim, o uso do nome '*Papai Noel*' só é possível porque há descrições disponíveis de Papai Noel, tal como '*o velhinho de barba branca e roupa vermelha que entrega presentes na noite de natal*'. São estas descrições que tornam inteligível o uso do nome, ou, dito de outra forma, que nos habilitam a identificar a sua referência. Um primeiro tipo de crítica a esta abordagem aponta uma suposta artificialidade na transformação de nomes em predicados. Um outro tipo de crítica, mais sério, foi primeiramente feita por Kripke (1980), que apresentou algumas situações de uso da linguagem que envolvem afirmações condicionais contrafactuais, em que tratar nomes como descrições leva a problemas insolúveis.

Deixando de lado este debate, podemos concluir que estas considerações de Quine complementam a teoria da denotação de Russell, que agora trata tanto das descrições impróprias quanto dos nomes impróprios. Podemos, portanto, ampliar nossa tabela de resumo acrescentando a avaliação que a abordagem Russell–Quine dá às sentenças ali presentes.

Sentença	**Frege**	**Russell–Quine**
O Atual rei da França é careca	I	F
Papai Noel não existe.	I	V
$\frac{1}{0} = \frac{1}{0}$	V	F

Para entender por que a abordagem Russell–Quine avalia como falsa a terceira sentença, $\frac{1}{0} = \frac{1}{0}$, basta notar que uma paráfrase sua em português seria '*O resultado da divisão de um por zero é idêntico a si próprio*', que, segundo a interpretação de Russell, afirma a conjunção (1) existe um único resultado da divisão de um por zero e (2) se ele existir é idêntico a si próprio. Como (1) é falsa, a sentença toda, conjunção de (1) com (2), também é falsa. Vale notar também a completa divergência entre a abordagem de Frege e a de Russell–Quine com relação às três sentenças da tabela.

A Abordagem da Tradição Matemática

Termos singulares indefinidos, que não denotam, tais como $\frac{1}{0}$, ou que denotam ambiguamente, tais como $\sqrt{4}$, são muito comuns na prática matemática. Uma das principais fontes de indefinição na matemática são os termos que representam a aplicação de uma função a um argumento para o qual ela não está definida.

O domínio de definição D_f de uma função f corresponde ao conjunto de valores para os quais ela está definida. Já o seu domínio de aplicação D^*_f corresponde ao conjunto de valores para os quais a função f pode ser aplicada em um dado contexto. Quando $D_f = D^*_f$, f é dita uma *função total* e nenhuma de suas aplicações será um termo indefinido. No entanto, quando $D_f \subset D^*_f$, f é dita uma *função parcial* e a aplicação $f(a)$ será um termo indefinido se $a \in D^*_f$ e $a \notin D_f$. Assim, se considerarmos por exemplo a função $\sqrt{\ } : \Re \rightarrow \Re$ (raiz quadrada com o conjunto dos números reais como domínio de aplicação e contradomínio), $\sqrt{\ }$ será uma função parcial pois seu domínio de aplicação é o conjunto dos números reais $D^*_f = \Re$, mas, restrita ao contradomínio dos reais, o seu domínio de definição são os números reais não negativos $D_f = \{x \in \Re \,/\, x \geq 0\}$, uma vez que não existe nenhum número real que multiplicado por si mesmo tenha como resultado um número negativo. Neste contexto, a expressão $\sqrt{-2}$ é um termo singular indefinido, pois teria como denotação um número que não existe. (FARMER, 2004, 475-476)

Poder-se-ia pensar que expressões como $\frac{1}{0}$ e $\sqrt{-2}$ são simplesmente erros, expressões mal escritas que ocorrem apenas nos cadernos de notas de estudantes, e que portanto, os termos indefinidos poderiam ser eliminados da boa prática matemática. Mas a situação não é tão simples assim. Considere, por exemplo, a seguinte expressão, também sugerida por Farmer (2004, 476):

$$\forall x \in \Re \ (x \geq 0 \Rightarrow (\sqrt{x})^2 = x)$$

Ela afirma que para todos os números x pertencentes ao conjunto dos números reais, se x for maior ou igual a 0 então o quadrado da raiz quadrada de x é igual a x. Nenhum matemático reclamaria que esta sentença não faz sentido, que ela está mal escrita ou envolve algum tipo de erro. Ela é uma sentença perfeitamente aceitável e verdadeira. No entanto, quando um lógico se depara com tal sentença e procura atribuir-lhe valor de verdade de acordo com a moderna teoria da satisfação de Tarski (1944), as coisas deixam de ser tão simples, pois para que uma sentença quantificada universalmente seja verdadeira, é preciso que todas as suas instâncias sejam verdadeiras, inclusive, neste caso, aquelas em que x é um número real negativo. Então a verdade incontestável de $\forall x \in \Re \ (x \geq 0 \Rightarrow (\sqrt{x})^2 = x)$ depende de nossa capacidade de avaliar e considerar verdadeiras sentenças como $(-2 \geq 0 \Rightarrow (\sqrt{-2})^2 = -2)$, que envolvem o termo singular indefinido $\sqrt{-2}$.

Este fato exemplifica que os termos singulares indefinidos ocorrem inevitavelmente na prática matemática e que não é possível eliminá-los simplesmente proibindo-os. Exemplifica também que a lógica clássica que, seguindo as prescrições de Frege, baniu de suas considerações os termos singulares indefinidos, ao exigir que todos os termos singulares tenham referência, não está totalmente adequada para formalizar todos os aspectos da prática matemática.

Farmer (2004), interessado em aperfeiçoar técnicas de formalização computacional da matemática, reconheceu o problema que os termos singulares indefinidos representam para este projeto, uma vez que tanto a lógica clássica de primeira ordem quanto a teoria simples dos tipos não admitem tais termos. Com vistas a propor um tratamento formal adequado ao modo como a tradição matemática tem lidado com os termos singulares indefinidos, Farmer compilou um conjunto de 3 princípios que sintetizam este tratamento. A lista abaixo, de 5 princípios, é apenas uma adaptação dos princípios compilados por Farmer à linguagem um tanto mais simples que temos adotado.

Princípios da Abordagem Matemática Tradicional aos Termos Singulares Indefinidos

1. Variáveis e constantes (termos atômicos) são sempre definidos. Sempre denotam.

2. Uma aplicação de função $f(a)$ pode ser indefinida, não denotar, se a não pertencer ao domínio de definição de f ($a \notin D_f$). Neste caso f não tem valor para o argumento a.

3. Uma aplicação $f(a)$ também pode ser indefinida se f relaciona mais de um valor possível ao argumento a. Neste caso f não é uma função, mas uma relação, não sendo, pois, um termo singular. Portanto, ao ser tomado como termo singular, sua denotação fica indefinida.

4. Fórmulas são sempre verdadeiras ou falsas, sempre definidas.

5. Uma fórmula atômica $\mathsf{P}(\mathsf{t})$ é considerada **falsa** se o termo t for indefinido.

Se, por exemplo, aplicarmos estes princípios à sentença $\frac{1}{0} = \frac{1}{0}$, vemos que em ambos os lados da identidade há um termo singular indefinido, $\frac{1}{0}$, justamente porque ele representa a aplicação de uma função, a divisão $div(x,y)$, a um par de argumentos $(1,0)$ para o qual ela não está definida.[11] Portanto, de acordo com o princípio 2, $\frac{1}{0}$ é um termo indefinido. Logo, de acordo com o princípio 5, como há termo singular indefinido na fórmula $\frac{1}{0} = \frac{1}{0}$, então ela é falsa segundo a abordagem tradicional.

Podemos, mais uma vez, ampliar nossa tabela de resumo das abordagens para incluir o tratamento tradicional da matemática.

Sentença	Frege	Russell–Quine	T. Matemática
O Atual rei da França é careca	I	F	$\times$
Papai Noel não existe.	I	V	$\times$
$\frac{1}{0} = \frac{1}{0}$	V	F	F

É digno de nota que, na única sentença para a qual a abordagem tradicional da matemática é pertinente, a terceira, ela concorda com a teoria de Russell–Quine e discorda da de Frege.[12]

A Abordagem Funcional

Pretendemos, nesta seção, propor uma nova interpretação semântica para as sentenças nas quais ocorrem nomes e descrições definidas, com o objetivo de poder ampliar para sentenças da linguagem comum a abordagem da tradição matemática para os termos singulares indefinidos.

[11] Seja div a função divisão tal que $div(m,n) = \frac{m}{n}$. Para perceber que $\frac{1}{0}$ é termo indefinido basta notar que o par $(1,0) \notin D_{div}$.

[12] Cabe ainda notar que a marcação '**I**' nas duas primeiras sentenças da coluna de Frege indica que a abordagem de Frege considera estas sentenças mas não atribui valor de verdade a elas, deixando-as indefinidas. Já a marcação '$\times$' nas duas primeiras colunas da abordagem da tradição matemática indica, por sua vez, que estas sentenças, por não serem matemáticas, não estão no escopo desta abordagem.

Um padrão semântico presente em todas as descrições definidas é que, conforme Frege já notara, todas elas podem ser reduzidas à forma '*o x tal que* $\varphi(x)$', onde $\varphi(x)$ representa uma propriedade expressa por um predicado ou fórmula. Além disso, em seus usos próprios, não problemáticos, para cada x existe um e apenas um indivíduo que satisfaz $\varphi(x)$.

No contexto da matemática, o caso mais fundamental de descrição definida se dá quando $\varphi(x)$ é uma fórmula atômica que afirma a identidade entre uma função aplicada ao argumento x e um valor especificado c. Ou seja, um caso padrão de '*o x tal que* $\varphi(x)$' é 'o x tal que $f(x)=c$'. Por exemplo, a descrição definida '*o número cujo cubo é oito*' é expressa na forma fregeana por '*o x tal que* $x^3=8$'. Neste caso, $\varphi(x)$ é expressa pela identidade $x^3=8$ que afirma que o valor da função cubo, $f(x)=x^3$, para o argumento x, é 8. Assim a expressão '*o x tal que* $x^3=8$' equivale a '*o argumento para o qual a função cubo tem valor 8*'. Mas uma maneira matematicamente mais direta de indicar qual é o argumento x para o qual a aplicação $f(x)$ tem a constante c como valor é através do conceito de função inversa (f^{-1}). De acordo com a definição padrão, f^{-1} é a função inversa de f quando $f(x)=y \iff f^{-1}(y)=x$. Se f usa o argumento x para denotar o valor y, f^{-1} denota x usando como argumento o valor y. Assim, a descrição '*o argumento para o qual a função f tem valor c*', ou seja, 'o x tal que $f(x)=c$' é equivalente a simplesmente $f^{-1}(c)$. Então, considerando que a função inversa da função cubo, $f(x)=x^3$, é a função raiz cúbica, $f^{-1}(y)=\sqrt[3]{y}$, a maneira matematicamente mais direta de denotar '*o x tal que* $x^3=8$' é dada simplesmente por '$\sqrt[3]{8}$'.

Nossa proposta aqui é usar este caso padrão de equivalência entre uma descrição definida e a aplicação de uma função

$$f^{-1}(c) \qquad \equiv \qquad \text{'o } x \text{ tal que } f(x)=c\text{'}$$

como modelo paradigmático para interpretar, através de funções, todos os casos de descrições definidas.

Vejamos como alguns exemplos de descrições definidas tanto na matemática quanto em linguagem natural podem ser traduzidos para aplicações de função:[13]

1) 'o autor de *Dom Casmurro*'
$\longrightarrow \quad oautorde(\mathsf{domcasmurro})$

2) 'o atual rei da França'
$\longrightarrow \quad oatualreide(\mathsf{franca})$

3) 'o número que multiplicado por 0 é 1'
$\longrightarrow \quad$ 'o x tal que $x.0=1$' $\quad \longrightarrow \quad \dfrac{1}{0}$

4) 'o número cujo quadrado é 4'
$\longrightarrow \quad$ 'o x tal que $x^2=4$ $\quad \longrightarrow \quad \sqrt{4}$

5) 'o elefante no guarda-roupa de Flammarion'
$\longrightarrow \quad oelefanteem(oguardaroupade(\mathsf{flammarion}))$

[13] Estamos, até aqui, adotando o padrão tradicional de notação de textos matemáticos semiformalizados, que consiste em usar fontes *itálicas* para variáveis, funções, predicados,... No entanto, quando escrevemos fórmulas na linguagem da lógica de primeira ordem, como os exemplos a seguir, estamos adotando um padrão mais adequado aos sistemas formais, no qual reservamos as fontes *itálicas* apenas para os símbolos de função. Para as variáveis e constantes individuais (nomes) usamos fontes sem serifa e apenas letras minúsculas, e para os símbolos de predicado usamos as mesmas fontes sem serifa, mas iniciando com letras Maiúsculas.

Nos exemplos (1) e (2) acima, a descrição definida pode ser traduzida diretamente para uma aplicação funcional. O termo 'o autor de *Dom Casmurro*' pode ser interpretado como a aplicação da função $oautorde$ ao argumento domcasmurro. A função $oautorde$ tem como domínio o conjunto de obras com autores individuais e como imagem o conjunto de pessoas que criaram estas obras. Assim, $oautorde(\mathsf{domcasmurro}) = \mathsf{machadodeassis}$, $oautorde(\mathsf{jubiaba}) = \mathsf{jorgeamado}$, e assim por diante. Da mesma forma, 'o atual rei da França' pode ser interpretado como a aplicação da função $oatualreide$ ao argumento franca. A função $oatualreide$ tem como domínio o conjunto de países que são monarquias governadas por reis e como imagem o conjunto destes reis. O termo 'o atual rei da França' é indefinido, não denota, porque a França é uma república. Não tem rei. Na notação funcional esta situação pode ser descrita em termos matemáticos ao dizermos que o termo $oatualreide(\mathsf{franca})$ é indefinido porque o argumento franca não está no domínio de definição da função $oatualreide$. Ou seja, $\mathsf{franca} \notin D_{oatualreide}$.

Em (3) e (4) temos novos exemplos do caso paradigmático que importamos da matemática discutido anteriormente, para o qual basta tomar a função inversa para obter a versão funcional da descrição definida. Nos dois exemplos temos termos singulares indefinidos. Em (3) porque o par $(1,0)$ não pertence ao domínio de definição da função $div(x,y) = \frac{x}{y}$, uma vez que não existe divisão por zero. E em (4) temos um termo indefinido porque $f(x) = \sqrt{x}$ não é uma função, já que tem mais de um valor para cada argumento, pois tanto 2 quanto -2 são $\sqrt{4}$.

O exemplo (5) mostra que mesmo casos mais complexos da linguagem natural podem perfeitamente bem ser parafraseados por aplicações de função. O termo 'o elefante no guarda-roupa de Flammarion' pode ser interpretado por um termo funcional que envolve duas aplicações de função. Vejamos como. A referência do termo (5) é um certo elefante. Esta referência é dada através da especificação de um certo lugar onde este elefante está: o guarda-roupa de Flammarion. Então uma primeira aplicação funcional toma a função $oelefanteem$, que tem como domínio o conjunto de lugares em que há elefantes e como imagem o conjunto destes elefantes, e aplica esta função a um certo lugar específico: o guarda-roupa de Flammarion. Mas este lugar, por sua vez, é também apresentado descritivamente através da função $oguardaroupade$, que tem como domínio o conjunto de pessoas que possuem um só guarda-roupa e como imagem estes guarda-roupas. Assim, o lugar onde está o elefante, argumento da função $oelefanteem$, é dado pela aplicação da função $oguardaroupade$ ao argumento flammarion, que é o nome de uma pessoa que supostamente tem um só guarda-roupa. Portanto podemos ver que a dupla aplicação de função $oelefanteem(oguardaroupade(\mathsf{flammarion}))$ interpreta adequadamente a descrição 'o elefante no guarda-roupa de Flammarion'.

Não encontrei ainda nenhum exemplo de descrição definida que não pudesse ser interpretada por aplicações de função nos moldes dos exemplos acima. Este modo funcional de interpretá-las tem ainda a vantagem de que a convenção matemática estabelece a *existência* e a *unicidade* da referência de $g(\mathsf{c})$ quando g é uma função total e $\mathsf{c} \in D_g$. Deste modo, podemos classificar os casos impróprios, aqueles em que há falha na denotação, nas categorias que a própria matemática nos dá. Assim, se $g(\mathsf{c})$ corresponde a uma descrição definida imprópria, um dos seguintes casos ocorre:

a) Há mais de um valor possível para a aplicação $g(\mathsf{c})$. Ou seja, g não é uma função, mas uma relação.

b) $\mathsf{c} \notin D_g$. A função g não está definida para o argumento c, que não pertence ao seu domínio de definição.

c) c é um termo indefinido. Ou seja, o nome c, ele próprio, não denota.

Esta lista é bastante conveniente, porque se qualquer termo indefinido enquadra-se em uma das três situações acima, então para propormos uma abordagem que trate de todos os casos impróprios possíveis, basta considerarmos estas três situações.

Uma vez que apresentamos uma interpretação funcional matematizada para as descrições definidas, nossa proposta é valermo-nos desta matematização e utilizar o tratamento que a tradição dá aos termos indefinidos, resumidos nos cinco princípios apresentados na seção anterior, para lidar com os casos problemáticos de falha na referencialidade. Estes princípios já nos indicam o tratamento dos itens (a) e (b) da lista acima, pois eles se enquadram exatamente nos princípios 5.3 e 5.2 respectivamente. O item (c), no entanto, não tem similar na tradição matemática, pois de acordo com o princípio 5.1 todas as variáveis e constantes deveriam denotar. Este caso exigirá um tratamento específico, que faremos mais adiante. Antes, vejamos alguns exemplos de proposições construídas com as interpretações funcionais de descrições definidas.

1) ‘O autor de *Dom Casmurro* é negro’
$\longrightarrow \quad \mathsf{EhNegro}(oautorde(\mathsf{domcasmurro}))$

2) ‘O atual rei da França viajou para Madri’
$\longrightarrow \quad \mathsf{ViajouPara}(oatualreide(\mathsf{franca}), \mathsf{madri})$

3) ‘O número que multiplicado por 0 é 1, é idêntico a si próprio’
$\longrightarrow \quad \frac{1}{0} = \frac{1}{0}$

4) ‘O número cujo quadrado é 4, é maior do que 0’
$\longrightarrow \quad \sqrt{4} > 0$

5) ‘O elefante no guarda-roupa de Flammarion está amarrotando sua camisa verde’
$\longrightarrow \quad \mathsf{EstaAmarrotando}(oelefanteem(oguardaroupade(\mathsf{fla})), acamisaverdede(\mathsf{fla}))$

6) ‘Papai Noel mora no polo norte’
$\longrightarrow \quad \mathsf{MoraPoloNorte}(\mathsf{papainoel})$

Vamos avaliar com as ferramentas que temos até o momento as interpretações funcionais das seis sentenças da lista acima. A sentença (1) é um caso simples e não problemático. A descrição definida ‘o autor de *Dom Casmurro*’ é interpretada pela aplicação de função $oautorde(\mathsf{domcasmurro})$ que denota um indivíduo específico, Machado de Assis, para o qual o predicado $\mathsf{EhNegro}$ se aplica. Portanto a sentença é verdadeira. Na sentença (2) a aplicação de função $oatualreide(\mathsf{franca})$ é um termo indefinido de acordo com o princípio 5.2, pois $\mathsf{franca} \notin D_{oatualreide}$. Assim, de acordo com o princípio 5.5 a sentença (2) é falsa. A sentença (3) também é falsa pelos mesmos motivos da (2), uma vez que nela ocorre o termo $\frac{1}{0}$ que é indefinido, também de acordo com o princípio 5.2, porque o par $(1{,}0) \notin D_{div}$. Na sentença (4) o termo $\sqrt{4}$ é indefinido, de acordo com o princípio 5.3, pois $\sqrt{x}$ tem mais de um valor para o argumento 4. Então, também pelo princípio 5.5, a sentença (4) é falsa. A sentença (5) pode ser verdadeira ou falsa, dependendo do modelo que se tome. O mais provável é que ela seja falsa e que sua falsidade se justifique pelo fato de não haver nenhum elefante no guarda-roupa de Flammarion. Neste caso teríamos um termo indefinido de acordo com o princípio 5.2 já que $oguardaroupade(\mathsf{fla}) \notin D_{oelefanteem}$. Por fim, ainda não temos condições de avaliar a sentença

(6) uma vez que ela envolve o uso de uma constante, papainoel, que, por não possuir denotação, é imprópria. Esta situação não está prevista nos princípios tradicionais da matemática, que exige que todas as constantes denotem.

Uma primeira solução possível para o problema imposto pela sentença (6) consiste, simplesmente, em adotar o tratamento predicativo que Quine dá aos nomes. Assim, o nome '*Papai Noel*' seria substituído pela descrição '*o indivíduo que é Papai Noel*' e a abordagem de Russell para esta descrição seria adotada. Então a sentença (6), 'Papai Noel mora no polo norte' seria interpretada exatamente da forma como a teoria Russell–Quine propõe:

$$\exists \mathsf{x}\,(\mathsf{EhPapaiNoel}(\mathsf{x}) \land \forall \mathsf{y}\,(\mathsf{EhPapaiNoel}(\mathsf{y}) \rightarrow \mathsf{x} = \mathsf{y}) \land \mathsf{MoraPoloNorte}(\mathsf{x}))$$

Mas esta solução nos parece insatisfatória, pois ela consiste em desistir de tratar o caso e adotar a abordagem Russell–Quine para o que não se conseguiu resolver. No lugar disso, vamos propor uma solução que tratará também os nomes como aplicações de função. Ou seja, vamos manter a consistência da abordagem, fazendo que tanto nomes quanto descrições definidas sejam tratados funcionalmente.

Propomos interpretar um nome como uma função identidade radicalmente parcial cujo domínio de definição ou é vazio, no caso de um nome impróprio, ou é um conjunto unitário, cujo único elemento é o objeto nomeado. Assim, o nome '*Jorge Amado*', por exemplo, ao invés de ser interpretado pela constante jorgeamado é interpretado pela função identidade parcial $oproprio\,jorgeamado$, cujo domínio é o conjunto unitário que tem Jorge Amado como o único elemento e, por ser uma função identidade, a imagem também é o próprio Jorge Amado. Então, quando o valor da variável x é Jorge Amado, a aplicação $oproprio\,jorgeamado(x)$ tem como valor (denota) o próprio Jorge Amado. Quando, por sua vez, o valor de x é distinto de Jorge Amado, então a aplicação $oproprio\,jorgeamado(x)$ é um termo singular indefinido, não denota, pois o argumento desta aplicação está fora do domínio de definição da função. No caso de um nome impróprio, que não denota, como '*Papai Noel*', o domínio da função $opropriopapainoel$ é vazio e portanto, todas as suas aplicações serão termos indefinidos. Ou seja, para qualquer valor de x, a aplicação $opropriopapainoel(x)$ será um termo singular indefinido que não denota. Em resumo, dado um nome c, o único valor possível para a função $oproprioc$ é o indivíduo que é c, e este valor só será atribuído quando a aplicação da função $oproprioc$ tiver como argumento o indivíduo que é c.

Assim, na interpretação da sentença (6), no lugar do nome papainoel usamos a função parcial identidade $opropriopapainoel$ e ficamos com algo como:

$$\mathsf{MoraPoloNorte}(opropriopapainoel(\mathsf{x}))$$

Como estamos agora introduzindo uma variável onde antes só havia constantes e funções, para podermos completar esta abordagem é preciso inserir esta variável no escopo de um quantificador. Afinal, a expressão acima não é uma sentença fechada, sobre a qual cabe julgar seu valor de verdade, mas é uma fórmula aberta cujo valor de verdade depende do valor de x. Mas quando simplesmente usamos um nome, sem mais esclarecimentos, este uso é compatível com a suposição da existência da entidade nomeada. Então é perfeitamente consistente completar a abordagem dos nomes como funções identidade parciais exigindo que as variáveis introduzidas nas substituições de nomes por aplicações de função estejam no escopo de um quantificador existencial. Assim, a completa interpretação da sentença (6), 'Papai Noel mora no polo norte' será:

$$\exists \mathsf{x}\, \mathsf{MoraPoloNorte}(opropriopapainoel(\mathsf{x}))$$

Podemos agora aplicar os princípios da tradição matemática a esta interpretação da sentença (6). Como 'Papai Noel' é um nome impróprio, o domínio de definição da função *opropriopapainoel* será o conjunto vazio. Portanto, para todos os valores de x a aplicação de função $opropriopapainoel(\mathsf{x})$ será, de acordo com o princípio 5.2, um termo impróprio, uma vez que se trata de aplicação de função a um argumento fora de seu domínio de definição. Assim, todas as instâncias de $\mathsf{MoraPoloNorte}(opropriopapainoel(\mathsf{x}))$ para todos os valores de x serão, de acordo com o princípio 5.5, falsas. Portanto, a sentença completa, $\exists \mathsf{x}\, \mathsf{MoraPoloNorte}(opropriopapainoel(\mathsf{x}))$ também será falsa.

Vamos, como um outro exemplo, reavaliar a sentença (1), levando em consideração nossa nova interpretação para os nomes. A nova interpretação para a sentença (1), 'O autor de *Dom Casmurro* é negro', fica:

$$\exists \mathsf{x}\, \mathsf{EhNegro}(oautorde(opropriodomcasmurro(\mathsf{x})))$$

Como '*Dom Casmurro*' é um nome próprio, com denotação, quando x denota este valor, a função $opropriodomcasmurro(\mathsf{x})$ é definida e também denota este valor, ou seja, denota a própria obra '*Dom Casmurro*'. Então, para este valor de x, a aplicação de função $oautorde(opropriodomcasmurro(\mathsf{x}))$ denota Machado de Assis, que por sua vez satisfaz o predicado $\mathsf{EhNegro}$. Portanto, como existe um valor de x que torna uma instância de $\mathsf{EhNegro}(oautorde(opropriodomcasmurro(\mathsf{x})))$ verdadeira, então a sentença (1) é, ela própria, verdadeira.

Podemos agora completar a tabela de resumo das abordagens aos termos indefinidos, incluindo esta nossa proposta de abordagem funcional.

Sentença	Frege	Russell–Quine	T. Matemática	A. Funcional
O Atual rei da França é careca	I	F	$\times$	F
Papai Noel não existe	I	V	$\times$	V
$\frac{1}{0} = \frac{1}{0}$	V	F	F	F

Não é difícil perceber que esta abordagem funcional concorda perfeitamente com a abordagem Russell–Quine na avaliação de todas as sentenças, sejam elas da matemática ou da linguagem natural, bastando, para tando, a análise de mais alguns exemplos. Para não nos alongarmos mais ainda, deixamos ao leitor que ainda não esteja convencido o ônus desta tarefa.

Considerações Finais

Propusemos a abordagem funcional aos nomes e descrições definidas com o intuito de ampliar para sentenças da linguagem natural a maneira tradicional da matemática abordar termos singulares indefinidos. Sugerimos, inicialmente, uma interpretação das descrições definidas como aplicações de função, utilizando como modelo o caso mais padronizado de descrição definida no contexto da matemática. Em seguida estendemos a abordagem para interpretar de modo funcional também os nomes, uma vez que no contexto da linguagem natural, estes termos também podem ser impróprios e não denotarem. Foi esta interpretação matematizada de nomes e descrições definidas que viabilizou a extensão da utilização dos princípios da tradição matemática para o tratamento de termos singulares indefinidos em linguagem natural; fora, portanto, do contexto matemático. O resultado desta extensão foi uma abordagem perfeitamente coincidente com a de Russell–Quine.

Esta coincidência evidencia que a abordagem Russell–Quine fundamenta-se sobre os mesmos princípios que embasam o modo matemático tradicional de lidar com termos indefinidos. (FARMER, 2004), de quem adaptamos os princípios apresentados na seção 5, não os compilou interessado em estudar os seus fundamentos filosóficos. Seu interesse, bastante prático, era o de identificar as regras lógicas deste tratamento de modo a propor alterações nas abordagens da lógica clássica de primeira ordem e teoria simples de tipos que o habilitassem a implementar em sistemas computacionais de matemática o mesmo tipo de abordagem a termos singulares indefinidos presente na tradição matemática. Por isso ele os propôs como regras práticas que indicam, em cada caso, a maneira tradicional de tratar os termos indefinidos.

Nosso interesse, ao contrário, está na comparação de abordagens diversas do tratamento de termos indefinidos e, portanto, volta-se para os fundamentos destas abordagens. Com isso em mente buscamos identificar um conjunto mais básico ainda de fundamentos que norteiam tanto os princípios da abordagem matemática tradicional, quanto da própria abordagem Russell–Quine. Podemos expressá-los da seguinte forma:

(i) Termos singulares indefinidos, que não denotam ou que denotam ambiguamente mais de um objeto, são perfeitamente legítimos e aceitáveis. Não são erros semânticos, não precisam ser evitados nem corrigidos.

(ii) Um termo singular indefinido não impõe a suspensão de julgamento sobre a verdade ou falsidade de uma proposição declarativa que o contenha. Proposições declarativas são, portanto, sempre verdadeiras ou falsas, independentemente de conterem ou não termos indefinidos.

(iii) O uso de um termo singular pressupõe a aceitação de sua referencialidade. A presença de um termo singular indefinido evidencia uma falha referencial que indica a frustração desta pressuposição. Portanto, uma proposição atômica que contenha algum termo singular indefinido será, por isso, falsa.

Podemos identificar o fundamento (i) tanto na abordagem Russell–Quine quanto na tradição matemática pelo fato de ambas aceitarem termos singulares indefinidos sem a imposição de nenhuma medida de proibição ou reforma. O fundamento (ii) também se faz notar no fato de que ambas abordagens valoram todas as proposições, não deixando a indefinição se propagar para o nível do julgamento sobre a verdade ou falsidade das proposições. Por fim, o fundamento (iii) está presente na abordagem Russell–Quine quando esta exige que se afirme a existência e unicidade dos termos individuais. Se alguma destas exigências falhar, falha também a referencialidade do termo e isto incidirá na falsidade da proposição. Este mesmo fundamento está presente explicitamente no princípio 5.5 da abordagem da tradição matemática.

É justamente porque ambas estas abordagens têm os mesmos fundamentos que a abordagem funcional coincide com a abordagem Russell–Quine. Afinal, propusemos a abordagem funcional apenas como uma ampliação do escopo da abordagem tradicional da matemática para o âmbito da linguagem natural. Obtida desta maneira por ampliação de escopo, ela também compartilha estes mesmos fundamentos e por isso coincide com a abordagem Russell–Quine.

E quanto às teorias de Frege e Meinong? Meinong parece simplesmente rejeitar que existam termos singulares indefinidos e, como consequência, rejeita os três fundamentos acima. Ele exige que todos os termos tenham referência individual e, por isso, todos os termos denotam. Já Frege, apesar de admitir a ocorrência de termos singulares indefinidos, principalmente no contexto da linguagem natural, não nutre nenhuma simpatia por eles, tratando-os como inconveniências linguísticas que necessitam de correção nos contextos mais controlados da matemática e da conceitografia. Ele, portanto, rejeita o fundamento (i) já que procura evitar e corrigir os termos indefinidos. Frege, além disso, rejeita igualmente o fundamento (ii), pois na versão Frege–Strawson de sua abordagem, que admite a ocorrência de termos singulares indefinidos, as sentenças nas quais estes termos ocorrem não são nem verdadeiras nem falsas. Ou seja, a presença de um termo singular indefinido impõe para Frege, contrariamente ao fundamento (ii), a suspensão sobre o julgamento de verdade ou falsidade. Quanto ao fundamento (iii), ainda que Frege não o rejeite explicitamente, sua abordagem o contradiz. De acordo com o fundamento (iii), qualquer proposição atômica em que ocorra um termo indefinido é falsa. No entanto, a versão Frege–Strawson considera indefinido o valor de verdade de uma proposição com termo singular indefinido e, nas versões Frege–Carnap e Frege–Grundgesetze, a atribuição artificial de referência aos termos indefinidos faz, por exemplo, da terceira proposição de nossa tabela comparativa ($\frac{1}{0} = \frac{1}{0}$) uma sentença verdadeira que, no entanto, a não ser por uma atribuição artificialmente forçada de denotação, possui termos singulares indefinidos. Sobre esta terceira proposição de nossa tabela, cabe ainda acrescentar que a abordagem de Frege, em todas as suas versões, discorda da abordagem tradicional da matemática que, juntamente com a abordagem Russell–Quine, a considera falsa.

Gostaria ainda de argumentar que a abordagem funcional que propusemos ajudou-nos a evidenciar a coincidência entre os fundamentos da abordagem Russell–Quine com os da tradição matemática e que, conforme pudemos ver, a abordagem de Frege é francamente oposta a estes fundamentos. Além disso, a abordagem funcional mostrou a plausibilidade e adequação da abordagem tradicional da matemática para o contexto mais amplo da linguagem natural. Considero que estes fatos dão suporte à abordagem Russell–Quine comparativamente à de Frege, por trazerem a tradição matemática como sua aliada.

Evidentemente o assunto não se encerra aqui. Há uma vasta e conhecida bibliografia que se desenvolveu desde então sobre o assunto, com muitas outras teorias que não tratamos e que se valem de ferramentas formais mais recentes e sofisticadas, como a semântica dos mundos possíveis, por exemplo. Entre elas podemos citar os trabalhos de (DONELLAN, 1966), (KRIPKE, 1980), (CHATEAUBRIAND, 2001), (EVANS, 1982), (NEALE, 1990), (SOAMES, 1998). Cabe, por isso, reforçar mais uma vez que a nossa intenção com a proposição de mais um abordagem foi primordialmente argumentativa, com o intuito de evidenciar a aproximação entre a teoria de Russell complementada pelas considerações de Quine e o tratamento tradicionalmente dado na prática matemática aos termos indefinidos. Tanto que o fato de nossa proposta de interpretação funcional, nos resultados de suas análises, coincidir com a teoria Russell–Quine, indica que não estamos defendendo uma nova posição teórica sobre a questão.

Finalizo com uma breve consideração sobre a lógica clássica e a sua relação com a matemática. A lógica clássica, certamente sob influência das ideias reformistas de Frege aqui expostas, baniu de sua consideração os termos indefinidos. Todos os nomes denotam e todas as suas funções são totais. Se considerarmos que a principal motivação para a consolidação da lógica clássica nos padrões que ela tem contemporaneamente foi a de constituir-se em um instrumento de formalização da matemática, esta sua característica de não admitir a ocorrência de termos singulares indefinidos, antes de ser uma qualidade é um defeito, uma vez que estes termos ocorrem inevitavelmente na prática matemática, conforme vimos na seção 5. A lógica clássica não é, portanto, a lógica da matemática. Esta sua rejeição a torna menos adequada para formalizar as inferências da prática matemática do que o são as lógicas livres, lógicas alternativas de um tipo que admite a ocorrência de termos que não denotam e que, por isso, conseguem formalizar a abordagem tradicional da matemática aos termos singulares indefinidos.

Referências

ALSTON, W. Ontological commitments. *Philosophical Studies*, v. 9, 1958.

BLACKBURN, S.; CODE, A. The Power of Russell's Criticism of Frege: 'On Denoting' pp. 48-50. *Analysis*, Oxford University Press on behalf of The Analysis Committee, v. 38, n. 2, p. pp. 65–77, 1978. ISSN 00032638. Disponível em: <http://www.jstor.org/stable/3327496>.

CARNAP, R. *Meaning and Necessity: A Study in Semantics and Modal Logic*. Chicago: University of Chicago Press, 1947. Disponível em: <http://www.bibsonomy.org/bibtex/27c5d4d13d322de4bbb4d2429717d7624/chb>.

CHATEAUBRIAND, O. *Logical Forms*. Campinas: UNICAMP, Centro de Lógica, Epistemologia e História da Ciência, 2001. v. 34. (Coleção CLE, v. 34). Part I - Truth and Descriptions.

CHATEAUBRIAND, O. Quine and ontology. *Principia*, v. 7, 2003.

CHURCH, A. *Introduction to Mathematical Logic (PMS-13)*. Princeton University Press, 1996. (Princeton Landmarks in Mathematics and Physics). ISBN 9780691029061. Disponível em: <http://books.google.com.br/books?id=JDLQOMKbdScC>.

DONELLAN, K. Reference and definite descriptions. *Philosophical Review*, v. 75, 1966.

DUMMETT, M. *Frege: Philosophy of Language*. Harvard University Press, 1993. ISBN 9780674319318. Disponível em: <http://books.google.com.br/books?id=EYP7uCZIRQYC>.

EVANS, G. *The Varieties of Reference*. Oxford: Oxford University Press, 1982.

FARMER, W. M. Formalizing Undefinedness Arising in Calculus. In: BASIN, D. A.; RUSINOWITCH, M. (Ed.). *IJCAR*. Springer, 2004. (Lecture Notes in Computer Science, v. 3097), p. 475–489. ISBN 3-540-22345-2. Disponível em: <http://dx.doi.org/10.1007/978-3-540-25984-8_35>.

FREGE, G. *The Basic Laws of Arithmetic: Exposition of the System*. University of California Press, 1964. Disponível em: <http://books.google.com.br/books?id=R1AVWnW1JvUC>.

FREGE, G. Sobre o Sentido e a Referência. In: *Lógica e Filosofia da Linguagem*. 2ª. ed. São Paulo: EdUSP, 2009. cap. 7, p. 129–158.

KAPLAN, D. What is Russell's theory of descriptions? In: D. F. PEARS. *Bertrand Russell: A Collection of Critical Essays*. New York: Anchor Books, 1972, (Modern studies in philosophy). p. 227–244.

KAPLAN, D. On the logic of demonstratives. *Journal of Philosophical Logic*, Springer Netherlands, v. 8, p. 81–98, 1979. ISSN 0022-3611. 10.1007/BF00258420. Disponível em: <http://dx.doi.org/10.1007/BF00258420>.

KRIPKE, S. *Naming and Necessity*. Oxford: Basil Blackwell, 1980.

LEWIS, D. *Counterfactuals*. [S.l.]: Basil Blackwell Ltd, 1973.

MEINONG, A. The Theory of Objects. In: CHISHOLM, R. M. (Ed.). *Realism and the Background of Phenomenology*. [S.l.]: Ridgeview Pub Co, 1981. ISBN 0917930142.

MENDELSOHN, R. *The Philosophy of Gottlob Frege*. Cambridge University Press, 2005. ISBN 9780521836692. Disponível em: <http://books.google.com.br/books?id=G6%5C_90xFwUbUC>.

NEALE, S. *Descriptions*. Ma: MIT Press, 1990.

PELLETIER, F. J.; LINSKY, B. Russell vs. Frege on definite descriptions as singular terms . In: N. GRIFFIN; D, JACQUETTE. *Russell vs. Meinong: the legacy of "On Denoting"* . New York: Taylor & Francis Routledge, 2009, (Routledge Studies in Twentieth-Century Philosophy , v. 29). p. 284. ISBN 9780415963640.

QUINE, W. V. O. On What There Is. In: *From a Logical Point of View*. New York: Harper Torchbooks, 1963. p. 1–19.

RUSSELL, B. On Denoting. *Mind*, v. 56, n. 14, p. 479–493, 1905.

RUSSELL, B. *The Philosophy of Logical Atomism*. [S.l.]: Routledge, 1985.

SEARLE, J. Russell's Objections to Frege's Theory of Sense and Reference. *Analysis*, Oxford University Press on behalf of The Analysis Committee, v. 18, n. 6, p. pp. 137–146, 1958. Disponível em: <http://www.jstor.org/stable/3326570>.

SMITH, J. F. The Russell-Meinong Debate. *Philosophy and Phenomenological Research*, International Phenomenological Society, v. 45, n. 3, p. pp. 305–350, 1985. ISSN 00318205. Disponível em: <http://www.jstor.org/stable/2107269>.

SOAMES, S. The modal argument: Wide scope and rigidified descriptions. *Noûs*, v. 32, 1998.

STRAWSON, P. On Referring. *Mind*, v. 59, p. 320–344, 1950.

STRAWSON, P. Singular terms and predication. *Journal of Philosophy*, v. 57, 1961.

TARSKI, A. The semantic conception of truth: and the foundations of semantics. *Philosophy and phenomenological research*, JSTOR, v. 4, n. 3, p. 341–376, 1944.

UMA VISÃO FILOSÓFICA ACERCA DA SEMÂNTICA ALGÉBRICA PARA A LÓGICA MODAL

Samir Gorsky*

Resumo: A lógica modal tem sido criticada por conta da sua ontologia subjacente, principalmente por Quine, um dos filósofos mais importantes a manter tal crítica. Seu principal argumento está fundado na ideia de que a semântica da lógica modal é baseada na noção de mundos possíveis e este conceito permite uma proliferação de entidades no universo (ver (QUINE, 1953)). A consequência disso é um grande problema em termos ontológicos. Entretanto, a semântica dos mundos possíveis não são a única semântica possível para a lógica modal; existem semânticas algébricas, por exemplo. Os trabalhos de Jósson, Mackinsey, Tarski e Lemmon nos anos quarenta, cinquenta e sessenta foram muito importantes para o desenvolvimento de tal semântica (ver (MCKINSEY, 1941a), (JóNSSON; TARSKI, 1951)). Bull e Segerberg em *Handbook of Philosophical Logic* (BULL; SEGERBERG, 1984) argumenta que se o artigo de Jósson e Tarski tivesse sido lido mais atentamente quando foi publicado, a história da lógica modal teria sido diferente. O esquema G^{mnpq} generaliza uma quantidade infinita de esquemas de axiomas (cf. (LEMMON; SCOTT; SEGERBERG, 1977)); baseado neste esquema, os resultados de completude e outras propriedades importantes podem ser provadas para uma quantidade infinita de sistemas modais em termos puramente algébricos. Estes resultados mostram que a semântica algébrica é tão apropriada quanto a semântica de mundos possíveis e, do ponto de vista filosófico, as semânticas algébricas podem ser vistas como uma resposta às críticas quineanas. Portanto, esses resultados são justificativas filosófica para o estudo da lógica modal e da sua semântica algébrica.

Palavras-chave: Lógica modal. Semântica algébrica.

Abstract: Modal logic has been criticized with respect to its underlying ontology, mainly by Quine, the most important philosopher who developed and maintained such criticism. His main argument is founded in the view that this semantics is based on the notion of possible worlds, and this concept allows a proliferation of entities in the universe (cf. (QUINE, 1953)). The consequence of this is a great problem in ontological terms. However, possible-world semantics are not the only possible semantics for modal logics: there are also the algebraic semantics, for instance. The work of Jósson, Mackinsey, Tarski and Lemmon in the years forty, fifty and sixties was very important for the development of such semantics (cf. (MCKINSEY, 1941a), (JóNSSON; TARSKI, 1951)). Bull and Segerberg in *Handbook of Philosophical Logic* (BULL; SEGERBERG, 1984) even claim that if the paper by Jósson and Tarski had been more widely read when it was published, the history of modal logic might have been different. The schema G^{mnpq} generalizes infinitely many modal axiom schemas (cf. (LEMMON; SCOTT; SEGERBERG, 1977)); based upon this schema, completeness results and other important properties can be proven for infinitely many systems of modal logics in purely algebraic terms. These results show that algebraic semantics is as appropriate as the possible-worlds semantics, and from the philosophical point of view algebraic semantics can thus be seen as a response to the Quinean criticisms. We have thus a philosophical justification for the study and development of both modal logic and its algebraic semantics.

Keywords: Modal logic. Algebraic semantics.

*Professor do Departamento de Filosofia da Universidade Federal do Rio Grande do Norte

Introdução

Começaremos este artigo com uma breve apresentação dos conceitos de semântica e completude para sistemas lógicos. Em seguida, abordaremos a temática da semântica algébrica. Este tipo de semântica é interessante quando relacionada às lógicas modais, pois é um exemplo de semântica que não faz referência a mundos possíveis. Apresentaremos um resultado de completude para as lógicas modais baseado nos artigos de Lemmon *Algebraic Semantics for Modal Logic I and II* (ALGEBRAIC..., 1966). Por fim, mostraremos que as semânticas algébricas para as lógicas modais podem ser consideradas adequadas uma vez que existe o teorema de representação para as álgebras modais.

Semântica

"A semântica de uma língua natural ou formal, é o conjunto de regras e princípios de acordo com os quais as expressões dessa língua são interpretadas" (BRANQUINHO, 2006). Desta forma, a semântica é uma área do conhecimento que tem por tema o significado dos termos de uma linguagem. Na definição acima também aparece o conceito de "interpretação". Esta interpretação consiste em estabelecer:

(1) O sentido das diversas expressões (simples ou compostas) de uma linguagem.

(2) A referências dessas mesmas expressões.

Os itens acima remetem ao problema já conhecido sobre referência e sentido (ver (FREGE, 1892)). Este problema é posto para as linguagem formais de maneira *sui generis*. A tarefa central da interpretação de uma linguagem formal é a construção do conceito de verdade para uma dada interpretação. A partir deste conceito de verdade em L (L é uma linguagem formal qualquer) é possível definir os conceitos restantes da semântica.

Definição 2.1. *Modelo: Uma interpretação I de L é um modelo de um conjunto* Γ *de fórmulas de L se, e somente se, todas as fórmulas de L resultam verdadeiras para I.*

Definição 2.2. *Consistência: Um conjunto Γ de fórmulas de L é consistente se, e somente se, Γ tem um modelo.*

Definição 2.3. *Fórmula logicamente válida: Uma fórmula φ de L é logicamente válida (denotação $\models_L \varphi$) se, e somente se, em todas as I que são modelos de Γ, φ é verdadeira.*

Definidos os elementos básicos de uma semântica para uma linguagem formal temos em mãos os instrumentos para a obtenção de resultados metateóricos. Como exemplos de tais resultados temos as provas de consistência e completude semântica para os sistemas formais. Em geral esses resultados fazem parte do que chamamos teoria de modelos.

Neste trabalho iremos nos concentrar nos resultados de completude dos sistemas modais. Grosso modo, dizemos que um sistema é completo se tudo o que queremos que seja teorema (ou seja, as fórmulas válidas ou verdadeiras) é de fato teorema. Segundo Church, a noção de completude, assim como a de consistência tem uma motivação semântica. A idéia é que os teoremas não devam ser conflitantes com as suas interpretações ((CHURCH, 1956) pg. 109).

Podemos dizer que um dos maiores projetos que encontramos ao estudar lógica é o de se construir um sistema, ou conjunto de sistemas, que contenha(m) todas as verdades da lógica pura. Isto ainda não foi feito, podemos então perguntar; e as verdades da lógica proposicional, já foram formalizadas em algum sistema? A resposta a essa pergunta é afirmativa e podemos encontrar várias provas diferentes para essa resposta (ver *Introduction to mathematical Logic* (CHURCH, 1956), *Mathematical Logic* (SHOENFIELD, 1973), *Model Theory* (KEISLER, 1992), *Introduction to Mathematical Logic* (MENDELSON, 1997) etc. para maiores informações sobre a lógica proposicional). Comecemos então a analisar o caso proposicional para então passarmos para a lógica modal, que é um exemplo de sistema estendido da lógica proposicional.

Em primeiro lugar, notemos que a linguagem da lógica proposicional é adequada para expressar qualquer função veritativa (ou função de verdade). Isto não significa que a linguagem da lógica proposicional é capaz de expressar qualquer verdade sobre funções veritativas (funções de verdade). A linguagem proposicional não pode, por exemplo, expressar a verdade de que existe um número infinito enumerável de distintas funções veritativas. Porém podemos provar (i) que a linguagem da lógica proposicional (que contenha como conectivos $\neg$ e $\rightarrow$) é adequada para expressar as funções veritativas (cf. (HUNTER, 1973) pp 62-67). A partir deste resultado podemos então mostrar que: (ii) qualquer fórmula verofuncional φ com conectivos arbitrários pode ser correlacionada a uma única fórmula na forma normal disjuntiva e com a mesma tabela de verdade. (iii) Qualquer fórmula na forma normal disjuntiva pode ser correlacionada a uma única fórmula da linguagem proposicional que tenha a mesma tabela de verdade que a primeira. Nada nos quarante, entretanto, que esta relação seja 1-1.

A completude será a idéia chave deste capítulo. Discutiremos um tipo especial de semântica (a semântica algébrica) e o resultado de completude desta semântica para certos sistemas modais.

Semântica algébrica

Ao analisarmos a história da lógica modal no século XX (ver capítulo 1 do presente trabalho) observamos que as lógicas modais foram estudadas primeiramente a partir de estruturas algébricas. Hugh MacColl foi o primeiro a fazer uma análise algébrica das proposições modais. Seu trabalho apareceu na revista *Mind* entre 1880 e 1906 em uma série de artigos sob o título *Symbolic Reasoning*(ver *Mathematical Modal Logic: a View of its Evolution* (GOLDBLATT, 2000) p. 4). A conjunção segundo MacColl era simbolizada por ab, e $a+b$ foi usado para simbolizar a disjunção. A implicação entre a e b (a implica b) tinha a seguinte forma: $a:b$ e a equivalência ficou $a=b$. A definição de equivalência era a mesma que conhecemos hoje em dia, ou seja, a composição de duas implicações $(a:b)(b:a)$. Para a negação MacCol usou o símbolo '. Assim, a' representava a negação de a.

Boole usou $a=1$ e $a=0$ para "a é verdadeiro" e "a é falso" respectivamente, porem deu um aspecto modal às suas definições por causa de sua leitura temporal destas (a é sempre verdadeiro e a é sempre falso) (ver *An Investigation of the Laws of Thought on Which are Founded the Mathematical Theories of Logic and Probabilities* (BOOLE, 1954) capítulo XI). MacColl usou as letras gregas ε e η para se referir aos conceitos de 'certeza' e 'impossibilidade'. Tais letras faziam o mesmo papel que o 0 e 1 do trabalho de Boole. Mais tarde MacColl ainda introduziu o símbolo θ para designar aquilo que não é nem certo nem impossível. Por fim, as expressões: $(a = \varepsilon)$, $(b = \eta)$ e $(c = \theta)$ expressavam 'a é certeza', 'b é impossível' e 'c é variável' (nem certeza, nem impossível). Tais fórmulas foram abreviadas por questão de simplicidade. Suas novas formas eram: a^{ε}, b^{η} e c^{θ}. Foram acrescentados ainda as seguintes fórmulas: d^{τ} para 'd é verdadeiro' e e^{ι} para 'e é falso'. Desta forma, a expressão $a : b$ era equivalente a $(ab')^{\eta}$ (é impossível que a e não-b).

Nos dias de hoje, temos visto um renovado interesse sobre a álgebra desenvolvida por MacColl. Por exemplo, o periódico *The Nordic Journal of Philosophical Logic* abriu espaço para o estudo dos trabalhos citados acima. Em particular o artigo *Hugh MacColl and the Algebra of Strict Implication* de Stephen Read (ver (READ, 1998)) apresenta uma argumentação sobre a interpretação do sistema apresentado por MacColl como sendo a lógica modal T desenvolvida mais tarde pelos trabalhos de Feys e von Wright (ver *Mathematical Modal Logic: a View of its Evolution (GOLDBLATT, 2000).*

Antes de começarmos o estudo das álgebras modais devemos distinguir os diferentes sistemas modais. A distinção entre sistemas modais pode ser feita a partir de matrizes. Um importante teorema acerca da caracterização dos sistemas modais a partir de matrizes se deve a Dugundji que em 1940 publicou seus resultados que mostravam a não possibilidade de termos matrizes finitas características para alguns sistemas modais. Existem ainda outras técnicas para a distinção dos sistemas modais. Duas das mais importantes destas técnicas são, o método algébrico empregado por McKinsey e Tarski (1941 e 1948) (ver *A Solution of the Decision Problem for the Lewis systems S2 and S4, with an Application to Topology* (MCKINSEY, 1941b) e *Some Theorems about the Sentential Calculi of Lewis and Heyting* (TARSKI, 1948a)) e o método semântico de Kripke (1959 e 1963) (ver *A completeness theorem in modal logic* (KRIPKE, 1959) e *Semantic analysis of modal logic I. Normal modal propositional calculi* (KRIPKE, 1963)). Os artigos escritos por Lemmon na década de 60 têm por objetivo apresentar uma síntese destes dois métodos. Um interessante resultado mostrado nestes artigos é que a completude semântica pode ser deduzida de resultados algébricos por meio de um teorema central. No primeiro artigo Lemmon tem por objetivo mostrar que método algébrico McKinsey-

Tarski, que é bem sucedido quando aplicado ao sistema $S4$, pode ser estendido para um grupo de seis sistemas modais onde o mais forte destes sistemas é o T. O segundo artigo trata de sistemas mais fortes e tinha-se como projeto um terceiro artigo que trataria da logica modal quantificada para todos os sistemas vistos (este terceiro artigo não chegou a ser escrito).

O método usado para trabalhar esses diferentes sistemas modais será o mesmo. Primeiro se estabelece uma correlação entre as matrizes regulares para cada sistema e um certo tipo de álgebra. Depois, usando as matrizes de Lindenbaum prova-se que cada sistema tem a propriedade do modelo finito e portanto que são decidíveis. Uma conseqüência disto é que poderemos restringir a nossa atenção para álgebras finitas. Por último será estabelecido o teorema de representação para cada sistema em termos de uma álgebra baseada no conjunto de todos os subconjuntos de um dado conjunto. Essas representações resultam na conexão entre o ponto de vista algébrico e o ponto de vista da semântica dos mundos possíveis (ou semântica de Kripke).

Sistemas modais fracos

Os sistemas modais fracos costumam não receber uma atenção especial dos filósofos. Talvez porque parece que tais sistemas não consigam dar conta da diversidade de interpretações filosóficas sobre as modalidades. Todavia, do ponto de vista da teoria geral da semântica, tais sistemas tem sido mais observados, pois é apenas através do estudo das lógicas modais fracas que podemos nos dar conta das limitações existentes com relação à semântica padrão atual (ver *Some remarks on (weakly) weak modal logics* (SCHOTCH, 1981)). Chamaremos de fraco (seguindo Lemmon, *Algebraic Semantics for Modal Logic I, (ALGEBRAIC..., 1966)* os sistema mais fracos ou equivalentes a T. Apresentamos a seguir os sistemas a serem analisados.

Axiomas:

Ax. 1 $A \to (B \to A)$

Ax. 2 $(A \to (B \to C)) \to ((A \to B) \to (A \to C))$

Ax. 3 $(\neg A \to \neg B) \to (B \to A)$

Ax. 4 $\Box(A \to B) \to (\Box A \to \Box B)$

Ax. 5 $\Box A \to \neg\Box\neg A$

Ax. 6 $\Box A \to A$

Regras:

Reg. 1 $A, A \to B \vdash B$

Reg. 2 $A \to B \vdash \Box A \to \Box B$

Reg. 3 $A \vdash \Box A$

Definições:

Df. 1 $A \vee B := \neg A \to B$

Df. 2 $A \wedge B := \neg(\neg A \vee \neg B)$

Df. 3 $A \leftrightarrow B := (A \to B) \wedge (B \to A)$

Df. 4 $\Diamond A := \neg \Box \neg A$

Df. 5 $A \Rightarrow B := \Box(A \to B)$

Df. 6 $A \Leftrightarrow B := (A \Rightarrow B) \wedge (B \Rightarrow B)$

Df. 7 $\Box^n A := \Box \underbrace{\ldots}_{'n' vezes} \Box A$

Df. 8 $\Diamond^n A := \Diamond \underbrace{\ldots}_{'n' vezes} \Diamond A$

Df. 9 $A \Rightarrow^n B := \Box^n(A \to B)$

Df. 10 $A \Leftrightarrow^n B := (A \Rightarrow^n B) \wedge (B \Rightarrow^n A)$

Os seis sistemas modais são definidos como se segue:

C2 = [A1 - A4 ; R1, R2]

D2 = [A1 - A5 ; R1, R2]

E2 = [A1 - A4, A6 ; R1, R2]

T(C) = [A1 - A4 ; R1, R3]

T(D) = [A1 - A5 ; R1, R3]

T = [A1 - A4, A6 ; R1, R3]

O cálculo proposicional é definido como [A1 - A3, R1] mais as definições D1 - D3. Portanto o cálculo proposicional clássico aparece como subestrutura dos sistemas modais apresentados. Vale ainda observar que em ambos os sistemas contendo A6 (E2 e T), A5 é dedutível e nos sistemas T(C), T(D) e T a regra R2 pode ser derivada. Portanto temos uma ordem de inclusão dos sistemas modais fracos , onde o sistema mais fraco pode ser visto como subsistema de um sistemas mais forte. Tal relação de ordem pode ser ilustrada pela seguinte tabela:

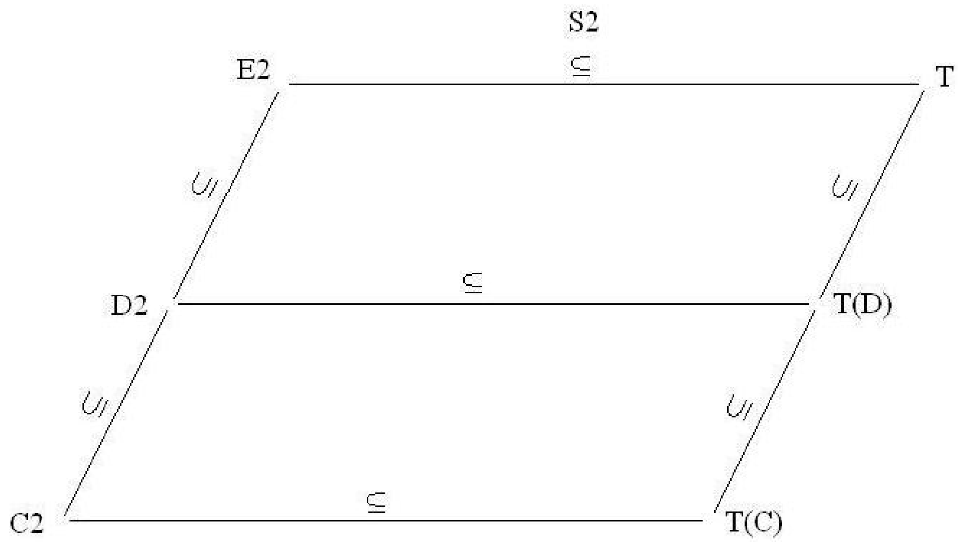


Figura 3: Tabela de inclusão dos sistemas modais fracos

Teorema 3.1. *A regra R2 é derivada nos sistemas T(C), T(D) e T.*

Proof. 1. $A \to B$ Pr.

2. $\Box(A \to B)$ 1, R3.

3. $\Box(A \to B) \to (\Box A \to \Box B)$ A4.

4. $\Box A \to \Box B$ 2,1 R1.

$\Box$

Primeiramente estudaremos o sistema C2. Podemos explicar tal decisão a partir do seu valor prático pois os resultados obtidos neste sistema podem facilmente ser estendido para os sistemas mais fortes. Por outro lado o sistema C2 é o sistema considerado minimal da lógica modal (embora alguns sistemas mais fracos possam ser encontrados).

Abaixo são apresentados alguns teoremas do sistema C2:

1) $\vdash_{C2} \Box(A \wedge B) \leftrightarrow (\Box A \wedge \Box B)$

2) $\vdash_{C2} \Diamond(A \vee B) \leftrightarrow (\Diamond A \vee \Diamond B)$

3) $\vdash_{C2} \Diamond(A \wedge B) \to (\Diamond A \wedge \Diamond B)$

4) $\vdash_{C2} (\Box A \vee \Box B) \to \Box(A \vee B)$

5) $\vdash_{C2} \Box^n A \to (B \Rightarrow^n A)$

6) $\vdash_{C2} \Box\neg A \to (A \Rightarrow^n B$

7) $\vdash_{C2} (A \Rightarrow^{n+m} B) \to (\Box^n A \Rightarrow^m \Box^n B)$

8) $\vdash_{C2} (A \Rightarrow^{n+m} B) \to (\Diamond^n A \Rightarrow^m \Diamond^n B)$

9) $\vdash_{C2} \Box^n A \leftrightarrow (\neg A \Rightarrow^n A)$

10) $\vdash_{C2} \Box^n A \leftrightarrow ((B \to B) \Rightarrow^n A)$

11) $\vdash_{C2} (A \Rightarrow^n B) \leftrightarrow \neg\Diamond^n(A \wedge \neg B)$

12) $\vdash_{C2} (A \Rightarrow^{n+m} B) \to ((B \Rightarrow^n C) \Rightarrow^m (A \Rightarrow^n C))$

Os seguintes teoremas de D2 e E2 não são teoremas de C2 (chamaremos estes resultados de (*)):

1) $\vdash_{D2} \Diamond(A \to)A)$

2) $\vdash_{D2} \Diamond(A \to A) \leftrightarrow (A \to A)$

3) $\vdash_{E2} A \to \Diamond A$

Um notável aspecto dos sistemas T(C), T(D) e T é resultado da regra R3 que prefixa o operador $\Box$ aos teoremas já demonstrados.

Teorema 3.2. *Se* $\vdash_{C2(D2,E2)} A$*, então* $\vdash_{T(C)(T(D),T)} \Box^n A$

Outro resultado interessante é o seguinte:

Teorema 3.3. $\vdash_T \Box(A \to A) \leftrightarrow (A \to A)$

Uma característica compartilhada por todos os seis sistemas visto mas que não aparece no sistema S2 é a possibilidade de substituição de equivalentes materiais no interior das fórmulas. O próximo teorema ilustra esta característica.

Teorema 3.4. *Para* $S \in \{C2, D2, E2, T(C), T(D), T\}$*. Se* $\vdash_S A \leftrightarrow B$*, então* $\vdash_S ...A... \leftrightarrow ...B...$ *(onde ...B... resulta de ...A... por substituição de zero ou mais ocorrências de A em ...A... por B)*

Proof. Prova: A prova se dá por indução no tamanho de ...A..., fazendo usos essenciais de R2 e R3.

$\Box$

Álgebras e Matrizes

Qual é a relação entre Álgebras e sistemas modais? Sabe-se que existe uma forte conexão entre o sistema de Lewis S4 e as álgebras-fecho (cf. (TARSKI, 1948b)). Uma conexão parecida acontece entre T e a generalização destas álgebras-fecho chamadas de álgebras-extensão (cf. (LEMMON, 1960) e (GOLDBLATT, 1976)). Mas será que podemos ter um método geral para determinar em cada sistema modal qual álgebra deve estar relacionada?

Para estudarmos o sistema C2 usaremos uma outra generalização de álgebra chamada simplesmente álgebra modal (Para maiores esclarecimentos sobre álgebra ver (SANKAPPANAVAR, 1981)).

Definição 3.5. *Uma estrutura* $\mathfrak{A} =< M, \cup, \cap, -, \boldsymbol{P} >$ *é uma* álgebra modal *sse M é um conjunto de elementos fechado sob as operações* $\cup, \cap, -$ *e* $\boldsymbol{P}$ *tais que:*

(i) M é uma álgebra Booleana com respeito a $\cup, \cap$ *e* $-$ *(Ver (SANKAPPANAVAR, 1981) para definição de álgebra booleana)*

(ii) Para $x, y \in M, \boldsymbol{P}(x \cup y) = \boldsymbol{P}x \cup \boldsymbol{P}y$

Definição 3.6. $\boldsymbol{N}x := -\boldsymbol{P} - x$

Teorema 3.7. *Em qualquer álgebra Modal* $\mathfrak{A} =< M, \cup, \cap, -, \boldsymbol{P} >$*:*

(i) Para $x, y \in M$*,* $\boldsymbol{N}(x \cap y) = \boldsymbol{N}x \cap \boldsymbol{N}y$

Esta é uma propriedade sobre a dualidade de P e N.

(ii) Para $x, y \in M$*, se* $x \leq y$*, então* $\boldsymbol{P}x \leq \boldsymbol{P}y$ *e* $\boldsymbol{N}x \leq \boldsymbol{N}y$ *(onde* $x \leq y$ *sse* $x \cup y = y$ *sse* $x \cap y = x$

Monotonicidade da relação $\leq$ *com relação a P e N.*

Proof. (i) $\mathbf{N}(x \cap y) = -\mathbf{P} - (x \cap y) = -\mathbf{P}(-x \cup -y) = -(\mathbf{P} - x \cup \mathbf{P} - y) = -\mathbf{P} - x \cap -\mathbf{P} - y = \mathbf{N}x \cap \mathbf{N}y$.

(ii) Suponha $x \leq y$, ou seja, $x \cup y = y$. Portanto $\mathbf{P}x \cup \mathbf{P}y = \mathbf{P}(x \cup y) = \mathbf{P}y$, daí $\mathbf{P}x \leq \mathbf{P}y$. Por outro lado se $x \leq y$, então $x \cap y = y$ e portanto $\mathbf{N}x \cap \mathbf{N}y = \mathbf{N}(x \cap y) = \mathbf{N}y$, logo $\mathbf{N}x \leq \mathbf{N}y$.

□

A seguir temos a apresentação da relação entre álgebras modais e matrizes.

Definição 3.8. *Dada uma assinatura C, uma C-matriz é um par* $M = < A, D >$, *onde* $A =< A, C >$ *é uma álgebra sobre C, e* $D \subseteq A$. *O conjunto D é normalmente referido como o conjunto de valores designados de M. As M-valorações de L(C) (L(C) é o conjunto de fórmulas da linguagem sobre C) são os C-homomorfismos* $v : L(C) \to A$.

Em geral uma álgebra pode ser transformada em uma matriz se algum subconjunto D de seus elementos é tomado como o conjunto dos elementos designados. Portanto para cada álgebra com uma cardinalidade N de elementos temos a possibilidade de construir 2^N matrizes distintas.

Uma lógica proposicional pode ser interpretada como matriz da seguinte maneira: Tomamos as variáveis proposicionais de uma fbf da lógica para ser a imagem sobre os elementos da matriz. Interpretamos os conectivos da lógica como operações na (ou definíveis na) matriz; deste modo, cada fbf A contendo n variáveis proposicionais está associada a uma (única) função-matriz $f^{(A)}$ de n variáveis.

Definição 3.9. *Dizemos que A é* satisfeita *por uma matriz sse, sob uma dada interpretação, o valor de* $f^{(A)}$ *para toda n-upla de elementos da matriz permanece em D. Caso contrário dizemos que A é* falsificada *pela matriz.*

Definição 3.10. *Um sistema S é* satisfeito *por uma matriz sse todos os teoremas de S são satisfeitos pela matriz.*

Definição 3.11. *Um sistema é* caracterizado *por uma matriz (ou uma matriz é* característica *para um sistema) sse as fbfs do sistema satisfeitas pela matriz são todas teoremas e apenas os teoremas deste sistemas.*

Recíprocamente, dada uma assinatura apropriada (ou seja, um conjunto de conectivos) e sua matriz-interpretação, temos que cada matriz determina uma lógica proposicional a saber: a lógica cujo os teoremas são exatamente a fbfs desta assinatura satisfeitas pela matriz sob a dada interpretação. Esta matriz também será a matriz característica para o sistema correspondente.

Mostramos a seguir como construímos uma matriz característica para uma dada lógica proposicional.

Para qualquer lógica proposicional L, seja W_L o conjunto de suas fbfs (em termo dos conectivos $c_1, c_2, ..., c_n$), e T_L o subconjunto de seus teoremas.

Teorema 3.12. *(Lindenbaum) Seja L uma lógica proposicional tal que T_L é fechada sob substituição de variáveis proposicionais. Então existe uma matriz característica $\mathfrak{M}$ para L.*

Proof. Notemos que a matriz não precisa ser finita. Suponha que L possua os conectivos $c_1^{a_1}, ..., c_n^{a_n}$, tais que $c_i^{a_i}$ é a_i-ádico ($1 \leq i \leq n$). Para os elementos da matriz tomaremos os membros de W_L e para os elementos designados os membros de T_L. Definimos, para cada $c_i^{a_i}$ uma função-matriz a_i-ádica c_i^* como se segue: para $w_1, ..., w_{a_i} \in W_L$ consideramos $c_i^*(w_1, ..., w_n)$ igual às fbfs que resultam da aplicação do conectivo $c_i^{a_i}$ às fbfs $w_1, ..., w_{a_i}$. Seja $\mathfrak{M}_L =< W_L, T_L, c_1^*, ..., c_n^* >$. Interpretando $c_i^{a_i}$ como c_i^*, é fácil ver que todos os teoremas de L são satisfeitos por $\mathfrak{M}_L$. De fato, suponha $t \in T_L$, então qualquer valoração dos argumentos de W_L para $f^{(T)}$ resulta como valor simplesmente a instanciação (por substituição) de t, que por hipótese está em T_L. Reciprocamente se $w \notin T_L$, então o valor de f^W está fora de T_L dado que estas valorações para estas variáveis consiste das variáveis proposicionais apropriadas de w. Logo $\mathfrak{M}_L$ é característica para L.

□

O teorema 3.3.12 nos mostra como conseguimos obter uma matriz característica a partir de uma lógica proposicional cujo o conjunto de teoremas é fechado para substituição de variáveis proposicionais. Como conseqüência temos o seguinte corolário.

Corolário 3.13. *Existe uma matriz característica para cada um dos sistemas C2, D2, etc.*

Apresentamos acima um método geral de se conseguir matrizes características para os sistemas. Todavia os resultados não nos indicam as matrizes que seriam mais interessantes para os nossos estudos. Usamos um método que resulta em matrizes de Lindenbaum. Um sistema pode ter muitas matrizes não isomórficas. A matriz mais interessante para trabalharmos a lógica proposicional é aquela que resulta da álgebra Booleana de dois elementos (ou seja, a matriz da tabela de verdade).

As matrizes que mais nos interessam neste ponto são estruturas $\mathfrak{M} =< M, D, \cup, \cap, -, \mathbf{P} >$ tal que $D \subseteq M$ e $\cup, \cap$ são operadores diádicos, $-, \mathbf{P}$ monádicos em M onde M é fechado para tais operações.

Definição 3.14. *Uma matriz* $\mathfrak{M} =< M, D, \cup, \cap, -, \boldsymbol{P} >$ *é* própria *sse* $D \subset M$.

Usaremos ainda as seguintes definições:

Definição 3.15. $x \to y := -x \cup y$

Definição 3.16. $x \leftrightarrow y := (x \to y) \cap (y \to x)$

Definição 3.17. *Uma matriz* $\mathfrak{M} =< M, D, \cup, \cap, -, \boldsymbol{P} >$ *é* regular *sse:*

(i) $\mathfrak{M}$ *é própria;*

(ii) D é ideal aditivo de M;

(iii) se $x \leftrightarrow y \in D$, *então* $x = y$.

(Um conjunto $D \subseteq M$ *é um* ideal aditivo *sse* $x \in D, y \in D \Rightarrow x \cap y \in D$ *e* $x \in D, y \in M \Rightarrow x \cup y \in D$.*)*

Os conectivos dos sistemas modais podem ser interpretados em termos de operações nas matrizes. Por exemplo $\Box$ é interpretado como **N**, os demais conectivos são interpretados da maneira usual, ou seja, $\vee$ como $\cup$, $\wedge$ como $\cap$, etc.

O próximo passo é mostrar como se constrói matrizes regulares a partir das matrizes de Lindenbaum. Vimos que existem matrizes de Lindenbaum para cada um dos seis sistemas estudados (C2, D2, E2, T(C) , T(D) e T). As matrizes regulares são construídas a partir de matrizes cujo os elementos são classes de equivalência sob o operador $\leftrightarrow$ das matrizes de Lindembaum. Partiremos então das seguintes definições:

Seja W o conjunto das fbfs dos seis sistemas estudados. Seja T o conjunto dos teoremas de C2 (D2, etc.). Para $A \in W$, considere $E(A) = \{B : B \leftrightarrow A \in T\}$. Seja $\vee, \wedge, \neg, \Diamond$ funções em W nas quais, para fbfs $A, B \in W$, formamos novas fbfs a saber: $A \vee B$,$A \wedge B$, $\neg B$ e $\Diamond A$. Definimos ainda novas funções $\vee_1, \wedge_1, \neg_1$ e $\Diamond_1$ sobre o conjunto de todos os conjuntos $E(A)$ da seguinte maneira:

$E(A) \vee_1 E(B) = E(A \vee B)$
$E(A) \wedge_1 E(B) = E(A \wedge B)$
$\neg_1 E(A) = E(\neg A)$
$\Diamond_1 A = E(\Diamond A)$

Chamamos de W_1 o conjunto de todos os conjuntos $E(A)$ para $A \in W$, e T_1 o conjunto de todos os conjuntos $E(A)$ para $A \in T$. Obviamente T_1 possui apenas um elemento, dado que, para $A, B \in T$, $A \leftrightarrow B \in T$.

Finalmente, definimos uma relação $\simeq$ sobre W de forma que $A \simeq B$ sse $A \leftrightarrow B \in T$ (mostrar que $\simeq$ é uma relação de congruência? isto é uma conseqüência do fato de que todos os seis sistemas possuem o cálculo proposicional como base, junto com o teorema 3.3.4 visto acima). Estamos, portanto, justificados em tratar a estrutura $\langle W_1, T_1, \vee_1, \neg_1, \Diamond_1 \rangle$ como uma matriz; chamamos esta matriz de $\mathfrak{M}_1$. Mostraremos abaixo que $\mathfrak{M}_1$ é uma matriz característica para C2 (D2, etc.).

Teorema 3.18. $\mathfrak{M} = \langle M, \{d\}, \cup, \cap, -, P \rangle$ *é uma C2-matriz regular sse* $\langle M, \cup, \cap, -, P \rangle$ *é uma álgebra modal e* $d = 1$.

Proof. Seja $\mathfrak{M} = \langle M, \{d\}, \cup, \cap, -, P\rangle$ uma C2-matriz regular. Para provar que $\langle M, \cup, \cap, -, P\rangle$ é uma Álgebra Booleana, precisamos tomar um conjunto de postulados para tais álgebras e mostrar que estes são satisfeitos pela matriz. (ver *The Elements of Mathematical Logic* (ROSENBLOOM, 1950), *A Course in Universal Algebra* (SANKAPPANAVAR, 1981) e *Algebraic Logic* (HALMOS, 1955)). Dos sete postulados apresentados por Rosenbloom as provas da satisfação de A1, A6, A7 pela matriz são conseqüências triviais do fato de $\mathfrak{M}$ ser uma matriz. Para A2, prova-se que $x \cap y = y \cap x$ como se segue: $(A \wedge B) \leftrightarrow (B \wedge A)$ é um teorema de C2, daí $x \cap y \leftrightarrow y \cup x \in \{d\}$ pois $\mathfrak{M}$ é uma C2-matriz. Segue-se ainda, de (iii) da Definição 5 que $x \cap y = y \cap x$. A3, A4, A5 são conseqüências similares do fato de que C2 contém o cálculo proposicional clássico, junto com (iii) da Definição 5. Assim $\langle M, \cup, \cap, -, P\rangle$ é uma álgebra modal. Visto que $\langle M, \cup, \cap, -\rangle$ é uma Álgebra Booleana, podemos considerar $1 = x \cup -x$. dado que $\vdash_{C2} A \vee -A$, temos $x \cup -x \in \{d\}$, ou seja, $x \cup -x = d$, daí $d = 1$.

Reciprocamente, seja $\mathfrak{M} = \langle M, \cup, \cap, -, P\rangle$ uma álgebra modal. Dado que $\langle M, \cup, \cap, -\rangle$ é uma álgebra Booleana, é óbvio que os esquemas A1-A3 são satisfeitos por $\langle M, \{1\}, \cup, \cap, -, P\rangle$. A função-matriz correspondente a A4 é $-N(-x \cup y) \cup (-Nx \cup Ny) = P(x \cap -y) \cup (P-x \cup -P-y) = P((x \cap -y) \cup -x) \cup -P-y = P(-x \cup -y) \cup -P-y = (P-x \cup P-y) \cup -P-y = P-x \cup 1 = 1$, portanto o esquema A4 é satisfeito. Para R1, suponha $x = 1$ e $x \to y = 1$. Então $y = (x \cap -x) \cup y = (x \cup y) \cap (-x \cup y) = (1 \cup y) \cap 1 = 1$. Para R2, suponha $x \to y = 1$. Então $-x \cup y = 1$ e $x \leq y$, daí $Nx \leq Ny$ pelo teorema 6 (ii). Assim $Nx \to Ny = -Nx \cup Ny = 1$. Assim $\langle M, \{1\}, \cup, \cap, -, P\rangle$ é uma C2-matriz. Que esta matriz é própria segue-se do fato que $\langle M, \cup, \cap, -\rangle$ é uma álgebra Booleana e portanto contém ao menos 2 elementos. É óbvio que $\{1\}$ é um ideal aditivo de M. Finalmente, dado $x \leftrightarrow y = 1$, $x = y$ pela propriedade encontrada nas álgebras Booleanas, logo $\langle M, \{1\}, \cup, \cap, -, P\rangle$ é também regular.

□

O teorema 3.3.18 nos indica como obtemos uma álgebra modal específica a partir de uma matriz regular e vice-versa (como obtemos uma matriz regular característica a partir da álgebra). Este teorema é uma generalização do teorema 4 que aparece em *An extension Algebras and the Modal System T* de Lemmon. O Teorema 3 do artigo de McKinsey *Solution of the Decision Problem for the Lewis systems S2 and S4, with an Application to Topology* tem este mesmo resultado para o sistema S2 e o teorema 10 deste mesmo artigo o faz para o sistema S4 (cf. (LEMMON, 1960) e (MCKINSEY, 1941b) p. 120).

Em virtude deste teorema não precisaremos mais distinguir a álgebra modal e a sua matriz correspondente que toma 1 como valor designado.

Para os demais sistemas devemos definir as álgebras apropriadas. Portanto:

Definição 3.19. *Uma álgebra é* deôntica *sse, além de ser modal, satisfaz o postulado:*

(iv) $P1 = 1$

Definição 3.20. *Uma álgebra é* epistêmica *sse, além de ser uma álgebra modal, satisfaz o postulado:*

(v) $x \leq Px$

Definição 3.21. *Uma álgebra é* normal *sse, além de ser uma álgebra modal, satisfaz o seguinte postulado:*

(vi) $P0 = 0$

Teorema 3.22. *Toda álgebra epistêmica é também uma álgebra deôntica.*

Proof. Por (v) nós concluímos $1 \leq P1$. Como se trata de álgebra booleana temos $P1 \leq 1$, daí $P1 = 1$, que é (iv).

□

O teorema acima assegura que os resultados obtidos para as álgebras epistêmicas podem ser inferidos a partir das álgebras deônticas.

Teorema 3.23. *A matriz $\mathfrak{M} = \langle M, \{1\}, \cup, \cap, -, P\rangle$ é uma D2- (E2-, T(C)-, T(D)-, T-) matriz regular sse $\langle M, \cup, \cap, -, P\rangle$ é uma álgebra deôntica (epistêmica, normal, normal deôntica, normal epistêmica) e $d = 1$.*

Proof. Em geral a prova se segue do teorema 3.3.18 acrescentado com as seguintes condições. Uma D2-matriz regular preenche a condição (iv) seguido do resultado (*) (2). Uma E2-matriz regular preenche a condição (v) seguido do resultado (*) (3). Uma T(C)-matriz preenche a condição (vi) seguido do teorema 4. Reciprocamente, dado (iv), $Nx \to Px = P - x \cup Px = P(-x \cup x) = P1 = 1$, desta forma o postulado A5 é satisfeito. Dado (v), $Nx \to x = P - x \cup x = (-x \cup P - x) \cup x$ (desde que $P - x = -x \cup P - x$) $= 1 \cup P - x = 1$, portanto o postulado A6 é satisfeito. Dado (vi), suponha $x = 1$. Então $Nx = -P - x = -P0 = -0 = 1$, daí R3 é satisfeito.

□

O resultado acima completa a relação entre álgebras e matrizes. Para cada álgebra é possível construir uma matriz regular adequada que identifica os teoremas do sistema. Como cada matriz está ligada a cada sistema lógico temos uma ponte entre a lógica e a álgebra. portanto, os resultados até aqui obtidos nos dão a completude para os seis sistemas modais da seguinte forma:

Teorema 3.24. $\vdash_{C2(D2,E2,T(C),T(D),T)}$ *A sse A é satisfeita por todas álgebras modais (deôntica, epistêmica, normal, normal deôntica, normal epistêmica).*

Proof. Em virtude do teorema 3.3.18 temos que C2 é satisfeito por qualquer álgebra modal; reciprocamente qualquer não-teorema de C2 é falsificado pela C2-matriz regular característica deste sistema (ver teorema 3.3.12), e assim as álgebras modais equacionam a fórmula correspondente a este não teorema ao valor 0 (teorema 3.3.18). Os outros sistemas são casos similares que usam os teoremas 3.3.23 e 3.3.12).

□

Portanto, podemos considerar o resultado acima como sendo um resultado de completude algébrica para as lógicas modais uma vez que mostra que a semântica algébrica "identifica" os (todos) teoremas de um dado sistema modal.

A propriedade dos modelos finitos

Mostraremos a seguir que os resultados obtidos acima não são alterados se restringirmos o seu campo de atuação para as álgebras finitas. Mostraremos ainda que cada sistema possui a propriedade do modelo finito, ou seja, cada não-teorema A de cada sistema pode ser associado a uma matriz finita satisfazendo o sistema e falsificando A. Daí poderemos também concluir que cada sistema é decidível.

Teorema 3.25. *Seja* $\mathfrak{M} = \langle M, \cup, \cap, -, P\rangle$ *uma álgebra modal (deôntica, etc.), e seja* $a_1, ..., a_r$ *uma seqüência de elementos de M. Então existe uma álgebra modal (deôntica, etc.) finita* $\mathfrak{M}_1 = \langle M_1, \cup_1, \cap_1, -_1, P_1\rangle$ *com no máximo* $2^{2^{r+1}}$ *elementos tais que:*

(i) Para $1 \leq i \leq r$, $a_i \in M_1$

(ii) Para $x, y \in M_1$, $x \cup_1 y = x \cup y$

(iii) Para $x, y \in M_1$, $x \cap_1 y = x \cap y$

(iv) Para $x \in M_1$, $-_1 x = -x$

(v) Para $x \in M_1$*, tal que* $Px \in M_1$, $P_1 x = Px$

Proof. Seja M_1 o conjunto de elementos de M obtidos de $P1, a_1, ..., a_r$ por qualquer número finito de aplicações de $\cup, \cap$ e $-$. Um resultado já conhecido da álgebra booleana nos indica que haverão não mais que $2^{2^{r+1}}$ elementos em M_1. Consideremos $\cup_1, \cap_1$, e $-_1$ iguais a $\cup, \cap$, $-$ mas restritos a M_1. Temos que é imediato o resultado para (i) a (iv). Para $x \in M_1$, dizemos que x é coberto por y sse $y \in M_1$, $Py \in M_1$ e $x \leq y$. Desde que $P1 \in M_1$ e também $1 \in M_1$ segue-se todo elemento é coberto por algum elemento. Se x é coberto por $y_1, ..., y_n$ nós consideramos $P_1 x = Py_1 \cap ... \cap Py_n$, (uma vez que $P_1 x \in M_1$). Se x está coberto por $y_1, ..., y_n$, então $x \leq y_i (1 \leq i \leq n)$, desta forma $Px \leq P_i$ pelo teorema 3.3.7 (ii) e $Px \leq P_1 x$. Reciprocamente, se $x \in M_1$ e $Px \in M_1$, então x é coberto por si mesmo, pois $P_1 x = Px \cap Py_1 \cap ... \cap Py_n$, onde $y_1, ..., y_n$ são os outros elementos que cobrem x. Assim $P_1 x \leq Px$. Estes resultados nos mostram que (v) é satisfeito.

Resta mostrar que se $\mathfrak{M}$ é uma álgebra modal, (deôntica, epistêmica, etc.), então $\mathfrak{M}_1 = \langle M_1, \cup_1, \cap_1, -_1, P_1 \rangle$ é também uma álgebra modal (deônti-ca, epistêmica, etc.). O resultado geral para álgebras modal depende da prova de $P_1(x \cup_1 y) = P_1 x \cup_1 P_1 y$. A prova de $P_1 x \cup_1 y = P_1 x \cup_1 P_1 y$ pode ser encontrada no artigo *Solution of the Decision Problem for the Lewis systems S2 and S4, with an Application to Topology*, (MCKINSEY, 1941b) p. 125. Para álgebras deônticas, precisamos mostrar que: dado $P1 = 1$, então $P_1 1 = 1$. Como $1, P1 \in M_1$, (v) se aplica e daí $P_1 1 = P1 = 1$. Para álgebras epistêmicas, precisamos mostrar que: dado $x \leq Px$, então $x \leq P_1 x$. Mas já mostramos que em geral $Px \leq P_1 x$. Para álgebras normais, precisamos mostrar que: dado $P0 = 0$, então $P_1 0 = 0$. Como, se $P0 = 0$, então $P0 \in M_1$, temos (por (v) novamente) $P_1 0 = P0 = 0$. Os demais casos são combinações dos casos já mostrados.

□

Teorema 3.26. *Seja A uma fbf com r subfórmulas. Então* $\vdash_{C2(D2,etc.)}$ *A sse A é satisfeita por todas álgebras modais (deôntica etc.) com, no máximo,* $2^{2^{r+1}}$ *elementos.*

Proof. Se $\vdash_{C2(D2,etc.)} A$, então, pelo teorema 3.3.24, A é satisfeito por todas álgebras modais (deônticas, etc.). Reciprocamente, suponha que A é um não-teorema de C2(D2, etc.) com r subfórmulas. Então A é falsificada com a matriz regular apropriada do teorema 3.3.12, a saber: $\mathfrak{M} = \langle M, \{1\}, \cup, \cap, -, P \rangle$. Sejam $v_1, ..., v_n$ as variáveis proposicionais em A e $a_1, ..., a_n$ os elementos de M que são associados a $v_1,...,v_n$ e que falsifica A. Suponha que, para esta função, os valores das subfórmulas de A diferentes de $v_1,...,v_n$ são $a{\cdot}n+1, ..., a_r$. Podemos assumir que última fórmula é a própria A, visto que: $a_r \neq 1$. Como $\langle M, \cup, \cap, -, P \rangle$ é uma álgebra modal (deôntica, etc.) (teoremas 3.3.18 e 3.3.23), portanto pelo teorema 3.3.25 existe uma álgebra modal (deôntica, etc.) $\mathfrak{M}_1$ com, no máximo, $2^{2^{r+1}}$ elementos satisfazendo a cinco condições daquele teorema. Podemos considerar (pela condição (i)) a mesma função que associa $a_1, ..., a_n$ a $v_1, ..., v_n$ na matriz $\mathfrak{M}_1$. É claro que esta função associa os mesmos valores para A (por (ii)-(v)) em $\mathfrak{M}_1$ que são associados em $\mathfrak{M}$ a saber: $a_r \neq 1$. Logo A é satisfeita por $\mathfrak{M}_1$.

□

Corolário 3.27. *Os sistemas C2, D2, E2, T(C), T(D), T possuem a propriedade do modelo finito, e são portanto decidíveis.*

Corolário 3.28. $\vdash_{C2(D2,etc.)} A$ *sse A é satisfeita por todas álgebras modais (deônticas, etc) finitas.*

Álgebras e modelos

Generalizaremos abaixo a noção de modelos kripkeanos. Mostraremos que, em termos de diferentes estruturas, os teoremas naturais de representação podem ser obtidos com relação às álgebras vistas acima. Dados estes teoremas, os resultados de completude de tipo kripkeano são conseqüências diretas dos resultados de completude da secção anterior.

Definição 3.29. *Um enquadramento modal bissortido (e.m.b) é uma tripla ordenada* $\mathfrak{R} = \langle K, Q, U \rangle$*, onde* K *é um conjunto não-vazio de elementos,* $Q \subseteq K$ *e* U *uma relação binária definida em* K*.*

Intuitivamente, podemos pensar K como um conjunto de "estados possíveis" e interpretar Uxy (Para $x, y \in K$) como "y é possivelmente acessível por x" ou "y é um mundo visível a x", etc.

Para entender o papel do subconjunto Q é necessário lembrar que os sistemas mais fracos C2, D2 e E2 não possuem teoremas da forma $\Box A$ e, portanto, que são consistentes mesmo se adicionarmos o esquema $\Diamond A$ como teorema (em pelo menos um mundo deverá valer $\neg A$ e, da mesma forma, em pelo menos um mundo deverá valer $\neg\neg A$ isto para qualquer A). De um ponto de vista interpretativo, isto significa que podemos permitir a possibilidade de mundos nos quais tudo é possível (incluindo contradições), Q portanto, é um conjunto de mundos com essas características (Este dispositivo foi sugerido por Lemmon em conversa com Saul Kripke como podemos constatar em (ALGEBRAIC. . . , 1966) p. 57).

Definição 3.30. *Dado um enquadramento modal bissortido $\mathfrak{R} = \langle K, Q, U\rangle$, definimos $\mathfrak{R}^+$ (a álgebra em $\mathfrak{R}$), como $\langle M, \cup, \cap, -, P\rangle$, onde:*

(i) $M = \wp K$

(ii) $\cup, \cap, -$ são como as operações (da teoria de conjuntos) união, intersecção e complemento restritas a M.

(iii) Para $A \in M$, $PA = \{x : \exists y(y \in A \wedge Uxy) \vee x \in Q\}$

Teorema 3.31. *Se $\mathfrak{R}$ é um enquadramento modal bissortido, então $\mathfrak{R}^+$ é uma álgebra modal.*

Proof. Que $\langle M, \cup, \cap, -\rangle$ é uma álgebra booleana é imediato da condição (ii) acima. Ainda, para $A, B \in \wp K$ temos:

$P(A \cup B) = \{x : \exists y(y \in A \cup B \wedge Uxy) \vee x \in Q\}$
$= \{x : \exists y(y \in A \wedge Uxy) \vee \exists y(y \in B \wedge Uxy) \vee x \in Q\}$
$= \{x : \exists y(y \in A \wedge Uxy) \vee x \in Q\} \cup \{x : \exists y(y \in B \wedge Uxy) \vee x \in Q\}$
$= PA \cup PB$

Portanto $\mathfrak{R}^+$ é uma álgebra modal.

□

Seja $\emptyset$ o conjunto vazio. Por (iii) é imediato:

Teorema 3.32. *Na álgebra de qualquer enquadramento modal bissortido:*

(i) $P\emptyset = Q$

(ii) $Q \subseteq PA$, para todo A.

Teorema 3.33. *(Jósson-Tarski) Qualquer álgebra modal finita é isomorfa à álgebra de algum enquadramento modal bissortido finita.*

Proof. Seja $\mathfrak{M} = \langle M, \cup, \cap, -, P\rangle$ uma álgebra modal finita. Então, pelo teorema de representação de Stone, $\langle M, \cup, \cap, -\rangle$ é isomórfico a álgebra de todos subconjuntos de um dado conjunto K. Seja φ este isomorfismo. Para $A \subseteq K$, tomemos $P'A = \varphi P \varphi_{-1} A$ e $Q = \varphi(P0)$. Para $x, y \in K, Uxy$ sse $x \in P'\{y\}$. Seja $\mathfrak{R} = \langle K, Q, U\rangle$, que é evidentemente uma estrutura modelo. Mostraremos que $\mathfrak{M}$ é isomórfica a $\mathfrak{R}+$ sob φ. Seja P^* a operação de possibilidade em $\mathfrak{R}_+$, ou seja, $P^*A = \{x : \exists y(y \in A0 \wedge Uxy\} \vee x \in Q\}$. É obvio que φ é um isomorfismo com respeito a $\cup, \cap$ e $-$, portanto resta somente mostrar que $\varphi(Px) = P^*(\varphi x)$.

Primeiro provaremos que para $x \in M, \varphi(Px) \cup Q = \varphi(Px)$. Usando o teorema 3.3.7 (ii), $P0 \leq Px$ daí $Px \cup P0 = Px$. Desta forma $\varphi(Px) = \varphi(Px \cup P0) = \varphi(Px) \cup \varphi(P0) = \varphi(Px) \cup Q$, por definição de Q. Agora, considere um átomo $a \in M$. Claramente φa é um conjunto unitário $\{u\}$ para algum $u \in K$, e portanto temos:

$$
\begin{aligned}
& P^*(\varphi(a)) = \{x : \exists y(y \in \{u\} \wedge Uxy) \vee x \in Q\} \\
& = \{x : Uxu \vee x \in Q\} \\
& = \{x : x \in P'u \vee x \in Q\} \\
& = \varphi(Pa) \cup Q \\
& = \varphi(Pa)
\end{aligned}
$$

Qualquer elemento $x \in M$ pode ser representado na forma $a_1 \cup ... \cup a_m$ para átomos $a_1, ..., a_m$. assim para $x \in M$:

$$
\begin{aligned}
& \varphi(Px) = \varphi(P(a_1 \cup ... \cup_m)) \\
& = \varphi(Pa_1 \cup ... \cup Pa_m \\
& = \varphi(Pa_1) \cup ... \cup \varphi(Pa_m) \\
& = P^*(\varphi(a_1)) \cup ... \cup P^*(\varphi(a_1)) \\
& = P^*(\varphi(a_1) \cup ... \cup \varphi(a_m)) \\
& = P^*(\varphi(a_1 \cup ... \cup a_m)) \\
& = P^*(\varphi(x))
\end{aligned}
$$

Logo, φ é também um isomorfismo com respeito a P.

$\square$

O teorema de Jósson-Tarski(outra versão): Qualquer álgebra booleana com operadores é imersível na álgebra complexa dos seus ultrafiltros (ver *Algebraic Tools for Modal Logic* (VENEMA, 2001)). Este teorema é semelhante em termos de resultado ao teorema abaixo.

O resultado acima serve como argumento para o posicionamento de Bull e Segerberg com relação à semântica algébrica para a lógica modal (posicionamento este mencionado na introdução do presente trabalho). De fato, ele mostra que a semântica algébrica possui características desejáveis como a decidibilidade. Isto acontece pois podemos focar a atenção sobre as álgebras modais finitas. Portanto os processos de "busca de valores" não se transformam em buscas infinitas. Porém o resultado acima é referente ao caso geral, isto é, à álgebra modal. Resta-nos estender este resultado para os outros sistemas.

Para os outros cinco tipos de álgebras, precisamos de definições apropriadas sobre as correspondentes estruturas-modelo.

Definiremos um enquadramento modal bissortido *deôntico* como sendo um enquadramento modal bissortido $\langle K, Q, U \rangle$ na qual U e Q satisfazem a condição:

(δ) Para $x \in K, \exists y Uxy$ ou $x \in Q$.

um enquadramento modal bissortido *epistêmico* é um e.m. $\langle K, Q, U \rangle$ na qual U e Q satisfazem a seguinte condição:

(ε) Para $x \in K, Uxx$ ou $x \in Q$.

((ε) é claramente equivalente à condição de que U seja reflexiva em $K - Q$).

Finalmente, definimos uma e.m. *normal* (e.m.n.) como uma e.m. $\langle K, Q, U \rangle$ na qual $Q = \emptyset$. Poderíamos pensar em e.m. de um e.m. normal como sendo uma estrutura $\langle K, U \rangle$ ao invés de uma estrutura $\langle K, Q, U \rangle$: onde não aparecem mundos "estranhos" (queer). Podemos reescrever a condição (iii) para a álgebra sobre um e.m. normal do seguinte modo:

Para $A \in \wp K$:

(iii)' $PA = \{x : \exists y(y \in A \wedge Uxy)\}$

De maneira similar, para uma álgebra sobre um e.m. normal deôntica, a condição (δ) pode aparecer na forma:

(δ)' para $x \in K, \exists y Uxy$

Para uma álgebra sobre um e.m. normal epistêmica, temos a condição que U seja reflexiva. Assim as e.m.'s normais epistêmicas são idênticas com o que Kripke chama de e.m.'s normais.

Teorema 3.34. *Se $\mathfrak{R}$ é uma estrutura modelo deôntica (epistêmica, normal, etc.), então $\mathfrak{R}^+$ é uma álgebra deôntica (epistêmica, normal, etc.).*

Proof. Seja $\mathfrak{R} = \langle K, Q, U \rangle$ um e.m. deôntica. Para se provar que $\mathfrak{R}^+$ é uma álgebra deôntica é suficiente, em virtude do teorema 3.3.31, provar que $PK = K$. Mas $PK = \{x : \exists y(y \in K \wedge Uxy) \vee x \in Q\} = K$ por condição (δ). Seja $\mathfrak{R} = \langle K, Q, U \rangle$ um e.m. epistêmica. Para se provar que $\mathfrak{R}^+$ é uma álgebra epistêmica é suficiente mostrar que (dado o teorema 3.3.31), para $A \subseteq K, A \subseteq PA$. Mas por (ε), para $x \in A, Uxx \vee x \in Q$, daí $\exists y(y \in A \wedge Uxy) \vee x \in Q$, portanto $x \in PA$. Seja agora $\mathfrak{R} = \langle K, Q, U$ um e.m. normal. Para provarmos que $\mathfrak{R}^+$ é uma álgebra modal devemos (teorema 3.3.31), provar que $P\emptyset = \emptyset$. Entretanto, por teorema 3.3.32 (i), $P\emptyset = Q$, mas $Q = \emptyset$. Logo, temos o resultado esperado.

□

O teorema acima, tal como está apresentado, foi demonstrado originalmente por Lemmon no artigo *Algebraic Semantics for Modal Logic I* (ver (ALGEBRAIC..., 1966) p. 59). É interessante notar que no artigo *An extension Algebras and the Modal System T* (LEMMON, 1960) Lemmon não chega a fazer o teorema de representação para T parando no teorema que mostra a obtenção de matrizes características para o referido sistema. Podemos então mostrar o teorema da representação.

Teorema 3.35. *Toda álgebra deôntica (epistêmica, normal, etc.) finita é isomórfica à álgebra de algum e.m. deôntica (epistêmica, normal, etc.) finita.*

Proof. Empregaremos a mesma terminologia e definições que foram usadas no teorema 3.3.33. A prova do isomorfismo é como a prova anterior. Resta, então, mostrar que $\mathfrak{R}$ é um e.m. deôntica (epistêmica, etc.). Primeiro suponha que $\mathfrak{M}$ é uma álgebra deôntica. Dado que $P1 = 1$, pelo isomorfismo $\varphi P^*K = K$ ou $\{x : \exists y Uxy \vee x \in Q\} = K\}$. A condição (δ) se segue imediatamente, e $\mathfrak{R}$ é portanto um e.m. deôntica. Segundo, suponha que $\mathfrak{M}$ é uma álgebra epistêmica. Como para $x \in M, x \leq Px$, pelo isomorfismo φ, para todo $A \subseteq A, A \subseteq P^*A$. Portanto, em particular para $x \in K, \{x\} \subseteq P^*\{x\}$, daí $x \in P^*\{x\}$. Temos então que $\exists y(y \in \{x\} \wedge Uxy) \vee x \in Q$, e (ε) segue-se imediatamente resultando no fato de que $\mathfrak{R}$ é um e.m. epistêmica. Terceiro: suponha $\mathfrak{M}$ uma álgebra normal. Daí $P0 = 0$ e pelo isomorfismo $\varphi(P^*\emptyset) = \emptyset$. Por definição $Q = \varphi(P0) = P^*(\varphi(0)) = P^*\emptyset$, logo $Q = \emptyset$ e $\mathfrak{R}$ é um e.m. normal.

□

Teorema 3.36. $\vdash_{C2(D2,etc.)} A$

(i) sse A é satisfeita por $\mathfrak{R}^+$ para toda e.m. $\mathfrak{R}$ (e.m. dôntica, etc.).

(ii) sse A é satisfeita por toda e.m. (e.m. deôntica, etc.) $\mathfrak{R}$ finita.

Proof. Se $\vdash_{C2} A$, então A é satisfeita por todas as álgebras modais (teorema 3.3.24), e portanto por $\mathfrak{R}^+$ para todas e.m. $\mathfrak{R}$ (teorema 3.3.31). Reciprocamente, se A é um não-teorema de C2, então existem álgebras modais finitas falsifica quando esta fórmula (teorema 3.3.26), e daí A é falsificada por $\mathfrak{R}^+$ para algum e.m. finita $\mathfrak{R}$ (teorema 3.3.33). Os outros casos são provados com emprego dos teoremas 3.3.34 e 3.3.35.

□

Falta-nos agora um resultado que mostre a equivalência desta completude para com a completude semântica (ao modo de Kripke). Conseguiremos por conseqüência direta do fato de que satisfatibilidade por $\mathfrak{R}^+$ para um e.m. $\mathfrak{R}$ é equivalente à validade em $\mathfrak{R}$ (no sentido kripkeano).

Em primeiro lugar, precisaremos de noções semânticas adequadas.

Definição 3.37. *Um* modelo *para uma fbf A em um e.m.* $\langle K, Q, U\rangle$ *é uma função binária* $\Phi(v,k)$*, onde v corresponde às variáveis proposicionais de A e K aos elementos de K, cujo os valores se encontram no conjunto* $\{T,F\}$.

A função definida acima consiste de uma valoração para cada variável proposicional de A em cada mundo de K. Assim, nesta função Φ cada variável está associada a um conjunto de mundos de K. Este conjunto indica quais os mundos onde a variável em questão possui valor T. Da mesma forma, cada mundo está associado a um conjunto de variáveis a saber: as variáveis que são verdadeiras neste mundo.

Definição 3.38. *Uma extensão única de* Φ, $\Phi'(B,k)$ *(onde B é uma subfórmula de A) é uma função definida como se segue:*

(i) Se B é atômica (isto é, uma variável), $\Phi'(B,k) = \Phi(B,k)$
(ii) $\Phi'(\neg C,k) = T$ *sse* $\Phi'(C,k) = F$
(iii) $\Phi'(C \to D,k) = T$ *sse* $\Phi'(C,k) = F$ *e/ou* $\Phi'(D,k) = T$
(iv) $\Phi'(\Box C,k) = T$ *sse* $\Phi'(C,l) = T$ *para todos* $l \in K$ *tal que* Ukl *e/ou* $k \notin Q$.

Como conseqüência desta definição, junto dom D1, D2 e D4, temos:

(v) $\Phi'(C \vee D,K) = T$ sse $\Phi'(C,l) = T$ para algum $l \in K$ e/ou $\Phi'(D,k) = T$
(vi) $\Phi'(C \wedge D,k) = T$ sse $\Phi'(C,k) = T$ e $\Phi'(D,k) = T$
(vii) $\Phi'(\Diamond C,k) = T$ sse $\Phi'(C,l)$ para algum $l \in K$ tal que Ukl e/ou $k \in Q$

Definição 3.39. *Dizemos que A é verdadeira em um modelo* $\Phi(v,k)$ *para* $l \in K$ *em um modelo* $\langle K, Q, U\rangle$ *sse* $\Phi'(A,l) = T$.

Definição 3.40. *Dizemos que A é válida em $\langle K, Q, U\rangle$ sse A é verdadeira em todos modelos $\Phi(v, k)$ para todos $l \in K$ em $\langle K, Q, U\rangle$.*

Definição 3.41. *Dizemos que A é válida sse A é válida em todas e.m..*

Definição 3.42. *Dizemos que A é D2- (E2, T(C)-, T(D)-, T-) válida sse A é válida em todas e.m. deônticas (epistêmicas, normal, normal deônticas, normal epistêmicas).*

Dada uma fbf A qualquer e um modelo $\Phi(v, k)$ para A em um e.m. $\mathfrak{R} = \langle K, Q, U\rangle$, podemos definir em termos de Φ uma valoração $\psi(\Phi)$ para as variáveis $v_1, ..., v_n$ de A a partir de $\mathfrak{R}^+$ pela seguintes definições: $V(v_i) = \{x : x \in K \wedge \Phi(v_i, x) = T\}(1 \leq i \leq n)$, e $\psi(\Phi) = \langle V(v_i), ..., V(v_n)\rangle$. Reciprocamente, dada uma valoração $\psi = \langle A_1, ..., A_n\rangle(A_i \subseteq K)$ a partir de $\mathfrak{R}^+$ para as n variáveis de A, podemos então definir um modelo $\Phi(\psi)(v, k)$?para A em $\mathfrak{R}$, tomando $\phi(\psi)(v_i, x) = T$ sse $x \in A_i$. Para qualquer valoração ψ de $\mathfrak{R}^+$ para as variáveis de A, e subfórmulas B de A, seja $V_\Psi(B)$ o valor associado para B em $\mathfrak{R}^+$ para a valoração ψ.

Lema 3.43. *(i) Seja A uma fbf e $\Phi(v, k)$ um modelo para A em $\mathfrak{R} = \langle K, Q, U\rangle$. Então para todo $x \in K\Phi'(A, x) = T$ sse $x \in V_{\psi(\Phi)}(A)$.*

(ii) Seja A uma fbf e ψ uma valoração para as variáveis de A a partir em $\mathfrak{R}^+$ para algum $\mathfrak{R} = \langle K, Q, U\rangle$. Então para todo $x \in K\Phi(\psi)'(A, x) = T$ sse $x \in V_\psi(A)$.

Proof. (i) Por indução no comprimento de A. Se A é uma das variáveis $v_1, ..., v_n$ o resultado vale por definição de $\psi(\Phi)$. Para o passo indutivo, suponha que o resultado vale para B e C. Primeiro notamos que pelos quantificadores lógicos:

$\forall y(Uxy \to \Phi'(B, y) = T) \leftrightarrow \forall y(Uxy \to y \in V_{\psi(\Phi)}(B))$

Agora suponha que A tem a forma $B \to C$. Então para todo $x \in K$:

$\Phi'(B \to C, x) = T \leftrightarrow \Phi'(B, x) = F \vee \Phi'(C, x) = T$

$\leftrightarrow x \notin V_{\psi(\Phi)}(B) \vee x \in V_\psi(\Phi)(C)$

$\leftrightarrow x \in (-V_{\psi(\Phi)}(B) \cup V_{\psi(\Phi)}(C))$
$\leftrightarrow x \in V_{\psi(\Phi)}(B \to C)$

Uma prova similar é obtida se A tem a forma $\neg B$. Suponha finalmente que A tem a forma $\Box B$. Então, para todo $x \in K$:

$\Phi'(\Box B, x) = T \leftrightarrow \forall y(Uxy \to \Phi'(B,y) = T \wedge x \notin Q$
$\leftrightarrow \forall y(Uxy \to y \in V_{\psi(\Phi)}(B)) \wedge x \notin Q$ (por (1))
$\leftrightarrow x \in NV_{\psi(\Phi)}(B)$
$\leftrightarrow x \in V_{\psi(\Phi)}(\Box B)$.

A prova de (ii) é similar: notemos que o resultado vale para o caso onde A é uma das variáveis $v_1, ..., v_n$ pela definição de $\Phi(\psi)$.

$\Box$

O principal teorema acerca da aproximação entre álgebra e semântica pode agora ser enunciado:

Teorema 3.44. *Seja $\mathfrak{R} = \langle K, Q, U \rangle$ um e.m., e A uma fbf qualquer. Então A é satisfeita por $\mathfrak{R}^+$ sse A é válida em $\mathfrak{R}$.*

Proof. Seja A uma fbf satisfeita por $\mathfrak{R}^+$, e considere um modelo $\Phi(v,k)$ para A em $\mathfrak{R}$. Então $V_{\psi(\Phi)}(A) = K$, daí pelo lema (i) $\Phi'(A,x) = T$ para todo $x \in K$. Assim A é válida em $\mathfrak{R}$. Reciprocamente, suponha que A seja válida em $\mathfrak{R}$ e considere uma valoração ψ para suas variáveis. Então, para todo $x \in K$, $\Phi(\psi)'(A,x) = T$, daí pelo lema (ii) para todo $x \in Kx \in V_\psi(A)$, portanto $V_\psi(A) = K$ e A é satisfeita por $\mathfrak{R}^+$.

$\Box$

Corolário 3.45. $\vdash_{C2(D2,E2,etc.)}$ *A sse A é válida (D2-válida, E2-válida, etc.).*

Proof. Usando-se o teorema 3.3.36 (i), o teorema 3.3.44 e as definições de validade obtemos o resultado de maneira direta. $\Box$

Conclusão

Para que os resultados acima fossem obtidos, o presente artigo apresentou e seguiu a metodologia aplicada por Lemmon (ALGEBRAIC. . . , 1966) a saber:

- **1º:** Estabelecer a correlação entre matrizes regulares para cada sistema e um certo tipo de álgebra
- **2º:** Usando a matriz de Lindenbaum, provar que cada sistema tem a propriedade do modelo finito.
- **3º:** Estabelecer os teoremas de representação para essas álgebras em termos de uma álgebra baseada em um conjunto de todos subconjuntos de um certo conjunto.

Além disso, é interessante indicar qual é o ponto de vista filosófico que pode ser desenvolvido sobre os resultados apresentados. O desenvolvimento da lógica modal não é menos interessante para a filosofia, mesmo se a semântica algébrica para a lógica modal tivessem sido inteiramente desenvolvida antes da semântica de Kripke. Todavia, provavelmente o desenvolvimento da lógica modal seria menos atratativo para os filósofos. Contudo, isso não implica que a lógica modal seria menos "filosófica".

Dentre as questões que vieram a motivar o presente trabalho, podemos destacar as seguinte:

- A semântica algébrica para a lógica modal é tão apropriada quanto a semântica dos mundos possíveis?
- Os filósofos contemporâneos teriam o mesmo interesse pela lógica modal se esta tivesse sido constituída como um sistema formal no qual a única semântica conhecida fosse esta que é construída a partir de uma classe de álgebras?
- A lógica modal teria se tornado uma área menor da matemática se sua semântica mais importante fosse algébrica?

A primeira questão foi respondida positivamente neste artigo. Dessa forma restam as duas outras questões. Essas são de um outro tipo, uma vez que sugerem uma especulação contrafactual sobre a semântica da lógica modal e não um resultado técnico verificável formalmente. É muito difícil saber quais seriam as posições detalhadas dos filósofos sobre a lógica modal caso a semântica dos mundos possíveis não tivesse sido proposta e verificada completa para os sistemas modais normais. Todavia, muito provavelmente, os filósofos da segunda metáde do século XX e início do século XXI teríam tido uma certa tendência para evitar argumentos baseados e/ou formalizados nesses sistemas se a única maneira de se atribuir significados a eles fosse a partir de certas classes de álgebras. A expressão "mundos possíveis", por si só, já possui um conteúdo filosófico notável. E anular a importância dessa semântica não é o objetivo de um resultado que tente mostrar a equivalência dessa semântica com a semântica algébrica.

Referências

ALGEBRAIC Semantics for Modal Logic I and II. *TheJournal of Symbolic Logic*, v. 31, n. 2, p. 46–65 and 191–218, 1966.

BOOLE, G. *An Investigation of the Laws of Thought on Which are Founded the Mathematical Theories of Logic and Probabilities*. [S.l.]: Macmillan, 1954.

BRANQUINHO, D. M. . N. G. G. J. *Enciclopédia de Termos Lógico-filosóficos*. São paulo: Martins Fontes, 2006.

BULL, R.; SEGERBERG, K. Basic modal logic. In: GABBAY, D.; GUENTHNER, F. (Ed.). *Handbook of Philosophical Logic*. [S.l.]: Springer, Dordrecht, 1984. p. 1–88.

CHURCH, A. *Introduction to mathematical Logic*. New Jersey: Princeton University Press, 1956.

FREGE, G. Über Sinn und Bedeutung. *Zeitschrift für Philosophie und philosophische Kritik*, p. 25–50, 1892.

GOLDBLATT, R. Metamathematics of modal logic. *Reports on Mathematical Logic*, v. 6, p. 41–77, 1976.

GOLDBLATT, R. *Mathematical modal logic: a view of its evolution*. 2000. Disponível em: <citeseer.ist.psu.edu/goldblatt01mathematical.html>.

HALMOS, P. *Algebraic Logic*. New York: Chelsea, 1955. v. 12. (Compositio Mathematica, v. 12).

HUNTER, G. *Metalogic, An Introduction to the Metatheory of Standard First Order Logic*. Berkley and Los Angeles: University of California Press, 1973.

JóNSSON, E.; TARSKI, A. Boolean algebras with operators. *American Journal of Mathematics*, v. 74, p. 891–939, 1951.

KEISLER, C. C. C. . H. J. *Model theory*. Amsterdam: North-Holland, 1992. v. 73. (Estudies in Logic and the fundations of matheamatics, v. 73).

KRIPKE, S. A completeness theorem in modal logic. *The Journal of Symbolic Logic*, v. 24, p. 1–14, 1959.

KRIPKE, S. Semantic analysis of modal logic i. normal modal propositional calculi. *Zeitschrift f?ur Mathematische Logik und Grundlagen der Mathematik*, v. 9, p. 67–93, 1963.

LEMMON, E. J. An extension algebras and the modal system t. *Notre Dame Journal of Formal Logic*, v. 1, n. 1, p. 3–12, 1960.

LEMMON, E. J.; SCOTT, D.; SEGERBERG, K. *The 'Lemmon notes' : an introduction to modal logic*. [S.l.]: Blackwell, 1977.

MCKINSEY, J. C. C. A solution of the decision problem for the lewis systems s2 and s4 with an application to topology. *Journal of Symbolic Logic*, v. 6, n. 4, p. 117–134, 1941.

MCKINSEY, J. C. C. A solution of the decision problem for the lewis systems s2 and s4, with an application to topology. *Journal of Symbolic Logic*, v. 6, n. 4, p. 117–134, 1941.

MENDELSON, E. *Introduction to Mathematical Logic*. London, Weinheim, New York, Tokio, Melburn, Madras: Chapman & Hall, 1997.

QUINE, W. V. O. *From a Logical Point of View*. [S.l.]: Harvard University Press, Cambridge, 1953.

READ, S. Hugh maccoll and the algebra of strict implication. *Nordic Journal of Philosophical Logic*, v. 3, n. 1, p. 59–83, 1998.

ROSENBLOOM, P. *The Elements of Mathematical Logic*. New York: Dover Publications, Inc., 1950.

SANKAPPANAVAR, S. N. B. . H. P. *A Course in Universal Algebra (millennium edition)*. Disponível para Download em http://onlinebooks.library.upenn.edu/webbin/book: [s.n.], 1981.

SCHOTCH, R. E. J. . P. K. Some remarks on (weakly) weak modal logics. *Notre Dame Journal of Formal Logic*, v. 22, n. 4, p. 309–314, 1981.

SHOENFIELD, J. R. *Mathematical Logic*. Massachusetts: Addison-Wesley, 1973.

TARSKI, J. C. C. M. . A. Some theorems about the sentential calculi of Lewis and Heyting. *jsl*, v. 13, n. 1, p. 1–15, 1948.

TARSKI, J. C. C. M. . A. Some theorems about the sentential calculi of lewis and heyting. *The Journal of Symbolic Logic*, v. 13, n. 3, p. 171–172, 1948.

VENEMA, M. G. . Y. Algebraic tools for modal logic. *ESSLLI 01*, 2001.

SOBRE AS RELAÇÕES DE CONSEQUÊNCIA LÓGICA E SEMÂNTICAS MULTIVALENTES

Carolina Blasio*

Resumo: A noção de consequência lógica criada por Tarski é definida como a preservação da verdade, ou de um conjunto de valores-de-verdade, das premissas para a conclusão. Uma vez que a noção de relação de consequência tarskiana é definida com base na bipartição dos valores-de-verdade da semântica, um dos resultados que segue de uma lógica tarskiana é que esta lógica pode ser caracterizada por uma semântica bivalente. Este resultado, conhecido como Redução de Suszko, implicaria a não existência de lógicas multivalentes, se não fosse a existência de noções de relações de consequências generalizadas associadas a semânticas cujo conjunto de valores possuem mais de duas partições. O objetivo do presente trabalho é contribuir para o debate acerca das diferentes noções de relação de consequência e a relação de algumas destas noções com as lógicas multivalentes.

Palavras-chave: Relação de Consequência Lógica, Semânticas Multivaloradas, Semânticas Multivalentes.

Abstract: The notion of logical consequence created by Tarski is defined as the preservation of truth, or a set of truth-values, from the premises to the conclusion. Since the notion of consequence relation is defined based on the bipartition of the semantics' truth-values, one of the results that follows from a Tarskian logic is this logic can be characterized by a bivalent semantics (Suszko's Reduction). This result would imply the absence of multivalent logics. Nonetheless, there are multivalent logics because there are generalized notions of consequence relations associated with semantics whose set of truth-values has more than two partitions. This paper aims to foment the debate about the several notions of consequence relation and the relationship of some of these notions with multivalent logics.

Keywords: Logical Consequence Relation, Manyvalued Semantics, Manyvalent Semantics.

*Doutoranda em Filosofia/UNICAMP. Este artigo foi elaborado com o apoio do CNPq (149706/2011-1).

Introdução

O conceito de relação de consequência lógica está na base da criação da Lógica como disciplina independente desde Aristóteles, sobretudo com as obras Analíticos Anteriores (*Αναλυτικων πσότεσων*) e Analíticos Posteriores (*Αναλυτικων υστεσων*). Seguindo a tradição antiga, dizemos que uma relação de consequência lógica trata de como enunciados (sentenças, proposições) seguem de outros e como se avalia este "seguir de".

Em linha gerais, existe uma relação de consequência quando um enunciado, a conclusão, segue de um conjunto de enunciados (que pode ser vazio), as premissas. Uma relação de consequência pode se dar na forma de um argumento, onde algumas proposições precisam ser escolhidas como premissas e outra como conclusão. Por exemplo,

$$(1)\quad \begin{array}{l} \text{Se chove, então o sertão floresce.} \\ \text{Chove.} \\ \hline \text{Logo, o sertão floresce.} \qquad \therefore \end{array}$$

No argumento (1), as premissas são os enunciados "se chove, então o sertão floresce" e "chove" e a conclusão é "o sertão floresce".

Há diferentes maneiras de definir como uma relação de consequência lógica se dá, pois esta depende da compreensão que se tem da própria Lógica. Uma das maneiras presentes na literatura para se definir uma relação de consequência é dada em termos da validade semântica, nela é dito que uma relação de consequência é válida porque a conclusão segue das premissas pela preservação da verdade, ou seja, uma relação de consequência é válida se, considerando que suas premissas sejam Verdadeiras, a conclusão deve também ser Verdadeira.

Uma das grandes questões que lógicos de diferentes áreas se colocam é sobre quais aspectos é definido uma relação como uma consequência lógica. Como ponto de partida, verificamos que as características mais tradicionais que espera-se de uma relação de consequência são a *necessidade*, a *formalidade* e a *normatividade* (Beal & Restall, 2006).

Geralmente o critério de validade de um argumento é definido como a verdade da conclusão seguir por *necessidade* da verdade de suas premissas. A necessidade é caracterizada de acordo com diferentes teorias metafísicas.

Sendo a necessidade um dos critérios para que uma relação de consequência se dê, apenas os argumentos ditos dedutivos são objetos da Lógica. Um exemplo de dedução seria:

(2)
Toda abelha é um artrópode.
Todo artrópode é um animal.
___
Logo, toda abelha é um animal. $\therefore$

Uma possibilidade de analisar o argumento (2) é considerarmos os termos "abelha", "artrópode" e "animal" como conjuntos e as duas premissas como Verdadeiras. Assim sendo, temos que o conjunto "abelha" é um subconjunto de "animal", logo a conclusão é Verdadeira. Em um argumento dedutivo, a conclusão segue *necessariamente* das premissas, pois, sendo as premissas Verdadeiras, a conclusão também será Verdadeira.

Existe, contudo, tipos de argumentos que não são dedutivos. Nestes argumentos, a verdade da conclusão é no máximo muito provável e não segue necessariamente da verdade das premissas. Charles Peirce (1974) identifica dois tipos de argumentos não dedutivos: a indução e a abdução.

Um argumento do tipo indutivo tem uma conclusão que se baseia no conteúdo dado pelas premissas. Por exemplo:

(3)
Toda criança que eu conheço gosta de brincar.
Maria é uma criança.
___
Logo, Maria gosta de brincar. $\therefore$

Diferentemente da dedução, em uma indução a conclusão pode não ser Verdadeira mesmo que as premissas sejam Verdadeiras.

Um argumento do tipo abdutivo tem uma conclusão que explica ou cria uma conjectura a respeito do conteúdo fornecido pelas premissas, por exemplo:

(4)
Observa-se que há mais míopes em relação ao passado.
Isto parece ocorrer porque vive-se mais em ambientes fechados.
___
Logo, viver em ambientes fechados é uma das causas da miopia. $\therefore$

Existem alguns estudos, principalmente em Filosofia da Ciência e Pragmatismo, que buscam formalizar o raciocínio indutivo e abdutivo. É reconhecido que tais tipos de raciocínio são essenciais para o desenvolvimento científico, entretanto, há algum debate sobre como formalizá-los e qual seria a legitimidade destes formalismos enquanto lógicas, uma vez estes que não contemplam o critério de necessidade. Há também quem defenda que a inviabilidade de formalizar tais argumentos e os incluem na chamada Lógica Informal.

A segunda característica tradicionalmente esperada da definição de relação de consequência é a *formalidade*. Em argumentos dedutivos uma relação de consequência pode ser unicamente determinada pela forma dos enunciados envolvidos, sem recorrer a algum tipo de conhecimento empírico. Podemos expressar os argumentos de forma esquemática, expondo a estrutura dos argumentos válidos. Por exemplo: "Se p então q, p; logo, q" é sempre um argumento válido para qualquer instância de p ou de q.

O argumento "Sofia é alérgica a camarão; logo, Sofia não pode comer casquinha de siri." não pode ser válido se sua formalização não levar em conta informações não declaradas, como "alergia a camarão provém do caso mais amplo de alergia a uma proteína presente em crustáceos", "o siri e o camarão são crustáceos", e também "uma pessoa pode morrer se ingerir um alimento ao qual ela tenha alergia".

A *normatividade* é a terceira característica esperada de uma relação de consequência lógica. A normatividade da relação de consequência, em geral, restringe que para um argumento ser válido, as premissas sejam Verdadeiras e a conclusão seja Falsa. Isto significa que um argumento sempre prescreve algo. Por exemplo: sabendo que o seguinte argumento é válido, "Se chover, Ana vai para casa, mas se fizer sol, Ana vai passear. Chove. Logo, Ana vai para casa.", podemos garantir que Ana vai para casa após verificar que chove e que a decisão de Ana era de ir para casa caso chovesse.

Os critérios para a noção de relação de consequência lógica podem ser expressos em termos da ideia wittgensteiniana de 'situação' (*Sachverhalte*). De acordo com Wittgenstein, um argumento válido é aquele cuja conclusão é Verdadeira em todas as situações em que todas as premissas são Verdadeiras (Wittgenstein, 1921). A relação de consequência lógica é definida de tal forma que é necessária, pois a verdade é preservada em todas as situações; é formal, pois se compromete com a estrutura, e não com o conteúdo do argumento; e também é normativa, pois não há um contraexemplo que demonstre que todas as premissas são Verdadeiras e a conclusão Falsa.

A relação de consequência tarskiana

A primeira definição formal da noção de relação de consequência entre os enunciados da Lógica Clássica de Primeira Ordem é atribuída a Rudolf Carnap pelo lógico polonês Alfred Tarski (1936): Uma sentença φ segue logicamente do conjunto de sentenças $\Gamma = \{\gamma_1, ..., \gamma_n\}$, $n \in \mathbb{N}$, se e somente se a união de Γ com $\neg\varphi$ é contraditória, ou seja, qualquer sentença segue de $\Gamma \cup \{\neg\varphi\}$. A definição de Carnap, contudo, não é suficiente para expressar uma sentença universal como consequência de infinitas instâncias individuais. Tarski redefiniu a noção de relação de consequência de Carnap em busca de uma noção formal ainda mais geral que fosse adequada ao que se espera de uma consequência lógica. Com a nova noção de relação de consequência Tarskiana torna-se possível demonstrar, por exemplo, que uma sentença universal da forma $\forall x\varphi$ é consequência de todas as instâncias de φ.

A noção de relação de consequência criada por Tarski é atualmente adotada com poucas modificações por grande parte dos sistemas lógicos formais. Para garantir que a noção de relação de consequência fosse geral e independente da linguagem lógica utilizada, Tarski empregou as definições semânticas de satisfação e modelo. Em termos gerais, dizemos que um certo objeto satisfaz uma determinada propriedade quando a atribuição desta propriedade ao objeto gera uma sentença verdadeira. Por exemplo, '(2, 3, 5)' satisfaz a função '$x + y = z$', ou '(João, Maria)' satisfaz o predicado 'x é irmão de y'.

Uma estrutura de interpretação possui um domínio, formado por objetos, e uma função que atribui um objeto do domínio a uma sentença. Dado um conjunto de sentenças Γ, dizemos que uma dada estrutura de interpretação é modelo de Γ, quando esta estrutura satisfaz todas a sentenças de Γ. O conceito de relação de consequência é então definido como: "Uma sentença φ é consequência semântica do conjunto de sentenças Γ se, e somente se, todo modelo de Γ é também modelo de φ". Esta é a definição da relação de consequência em termos de *teoria de modelos*, onde a validade ocorre pela preservação da verdade, ou seja, quando, dada todas premissas Verdadeiras, a conclusão também é Verdadeira. Ou ainda, a validade ocorre quando é incompatível que as premissas sejam Verdadeiras e a conclusão seja Falsa.

Uma relação de consequência tarskiana respeita as seguintes propriedades:

Reflexividade: Um enunciado segue de si mesmo.

Monotonicidade: Se um enunciado φ segue de um conjunto de enunciados Γ, então este mesmo enunciado φ segue de uma extensão deste conjunto de enunciados Γ', onde Γ é subconjunto deΓ'.

Transitividade: Se um enunciado φ segue de um conjunto de enunciados Δ e todo enunciado $\delta \in \Delta$ segue de um conjunto de enunciados Γ, então φ segue de Γ.

Compacidade: Se um enunciado φ segue de um conjunto de enunciados Γ, então existe um subconjunto finito Γ' de Γ, tal que φ segue de Γ'.

Estruturalidade: Seja φ um enunciado que segue de um conjunto de enunciados Γ. Então, se todas ocorrências de uma variável que ocorre em φ e nos enunciados de Γ forem substituídas de tal forma que φ^* e Γ^* sejam as enunciados resultantes desta substituição, então φ^* segue de Γ^*.

Outras noções de relação de consequência

Apesar da noção de relação de consequência tarskiana ser amplamente adotada, notamos na literatura da Lógica algumas noções de relação de consequência diferentes, sobretudo, nas lógicas não-clássicas. Categorizamos algumas destas diferentes noções em três tipos em comparação com a noção de relação de consequência tarskiana: as *extensões*, as *dissidências* e as *generalizações*. Note que estas categorias não são excludentes entre si e nem tem o propósito de exaurir todas as noções de relação de consequência existentes.

Consideramos como uma *extensão* da noção de relação de consequência tarskiana a relação de consequência de conclusão múltipla desenvolvida por Shoesmith & Smiley (1978). A noção de relação de consequência de conclusão múltipla tem sua raiz no cálculo de sequentes de Gentzen criado na década de 1930. A relação de consequência de conclusão múltipla estende conservativamente a noção tarskiana de relação de consequência admitindo que os argumentos possam ter múltiplos enunciados como conclusão. Um argumento teria, portanto, um conjunto de enunciados como premissas, intuitivamente compreendido como uma conjunção de enunciados, que acarretaria um conjunto de enunciados alternativos como conclusão, intuitivamente compreendido como uma disjunção de enunciados. Por exemplo: "Se chove, então o sertão se enche de flor; *e* chove; logo, o sertão se enche de flor; *ou* irá fazer calor no sertão".

A relação de consequência de conclusão múltipla possui tanto opositores quanto defensores: Por um lado, filósofos antirrealistas questionam o uso da conclusão múltipla por alegar que ela é pouco intuitiva, pois a conclusão, ao ter vários enunciados alternativos, não aponta para qual destes enunciados é, de fato, consequência das premissas (Rumfitt, 2008; Dummett, 1991).

Por outro lado, a noção de consequência de conclusão múltipla é elegante por sua simetria e maior expressividade. Um argumento é válido quando, sendo todas as premissas Verdadeiras, algum enunciado da conclusão é Verdadeiro; ou ainda, um argumento é válido quando, sendo todos enunciados da conclusão Falsos, alguma das premissas é Falsa. Outra forma de expressar a validade é pela incompatibilidade de que todas premissas sejam Verdadeiras e todas as alternativas da conclusão sejam Falsas (Beal & Restall, 2006).

Uma das vantagens de se empregar a relação de consequência de conclusão múltipla é sua maior expressabilidade. Por exemplo, o principio de explosão da lógica clássica é geralmente expresso pela noção tarskiana como $\alpha \wedge \neg\alpha \vdash \beta$ (ou seja, é incompatível que $\alpha \wedge \neg\alpha$ seja Verdadeiro e β seja Falso). Com a noção de conclusão múltipla o princípio de explosão é expresso com maior economia de linguagem: $\alpha, \neg\alpha \vdash^m$ (ou seja, é incompatível que α e $\neg\alpha$ sejam Verdadeiros.).

A noção tarskiana de relação de consequência também possui a limitação de não diferenciar certas lógicas. Lógicas tarskianas cujos argumentos são sempre válidos não são diferenciáveis das lógicas em que todos os enunciados de sua linguagem são verdadeiros. E ainda, a relação de consequência tarskiana não diferencia lógicas em que todos os enunciados de sua linguagem são falsos das lógicas em que todos os enunciados de sua linguagem possuem um único e mesmo valor. Estas distinções são expostas quando a relação de consequência de conclusão múltipla é adotada (Marcos, 2005).

O grupo de noções *dissidentes* de relações de consequência lógica é caracterizado pela modificação da definição de relação de consequência tarskiana com o propósito de fornecer uma alternativa. Neste grupo destacam-se a relação de consequência das lógicas relevantes e a das lógicas intuicionistas, ou construtivistas. Estas noções, em geral, possuem propriedades semelhantes à noção de relação de consequência tarskiana, mas discordam de seus fundamentos conceituais. Em geral, as relações de consequência dissidentes estão no cerne dos debates entre a existência de um único tipo de lógica, o monismo lógico, e a posição de que haveriam noções igualmente boas de relações de consequência, o pluralismo lógico.

A relação de consequência das lógicas *relevantes* restringe a compreensão do que seja "seguir de". Falha o princípio de explosão: um enunciado qualquer β não segue de $\alpha \wedge \neg\alpha$, uma vez que β não está implícito em $\alpha \wedge \neg\alpha$. Além disso, também falha o princípio do terceiro excluído: nem sempre um argumento cuja conclusão seja da forma $\alpha \vee \neg\alpha$ é válido. Não basta que a conclusão seja Verdadeira em todos os casos que as premissas sejam Verdadeiras, pois as premissas precisam também ser relevantes para a conclusão. A noção de consequência relevante exige que as premissas e a conclusão possuam termos ou (sub)fórmulas em comum. Exemplo: "Natal é uma cidade ensolarada. Ponta Negra é uma praia urbana de Natal. Os turistas gostam de Ponta Negra. Logo, os turistas gostam de uma praia urbana de Natal", pois os termos "turistas" e "praia urbana de Natal" presentes na conclusão fazem parte das premissas. Não é possível concluir, no entanto, que "ou os turistas gostam de Pipa ou os turistas não gostam de Pipa", pois "Pipa" não é um termo das premissas.

A noção de consequência das lógicas *intuicionistas*, ou construtivistas, exige que a partir das premissas se possa apresentar uma construção da conclusão. Esta noção não admite a lei do terceiro excluído ($\vdash \varphi \vee \neg\varphi$) nem a regra da eliminação da dupla negação ($\neg\neg\varphi \vdash \varphi$). Para um argumento ser válido, dadas premissas Verdadeiras e justificadas por uma construção de acordo com certas regras, a conclusão deve ser Verdadeira e justificada por uma construção de acordo com as mesmas regras usadas na construção das premissas.

As noções de consequência lógica que *generalizam* a noção tarskiana tem como característica não possuir uma ou mais de suas propriedades. As mais conhecidas e estudadas destas noções são as relações de consequência não-monotônicas, em que falha a propriedade de monotonicidade. A noção de consequência não-monotônica foi desenvolvida para capturar e representar inferências que representam melhor o nosso cotidiano, onde existem situações em que algo é concluído de forma válida e situações em que a mesma coisa é invalida devido a alguma nova informação. Por exemplo: "Todo pássaro voa. Piu-piu é um pássaro. Logo, Piu-piu voa" é um argumento válido, mas ao acrescentarmos "pinguins não voam" e "Piu-piu é um pinguim" como premissas, o argumento deixa de ser válido.

Outras classes de relações de consequência generalizadas apareceram recentemente na literatura e possuem importância no estudo das lógicas não-clássicas multivaloradas. A primeira a surgir foi a *quasi-consequence*, ou emph*q*-consequence de (Malinowski, 1990), que não é reflexiva. A seguir surge outras definições como a *plausible-consequence*, ou *p-consequence*, de

(Frankowski, 2004), que não é transitiva. Como veremos a seguir, estas relações colocam em xeque a própria noção de relação de consequência ao não possuírem propriedades consideradas fundamentais da noção de consequência tarskiana.

Multivaloração e multivalência

Vimos que a noção de consequência tarskiana tem como característica a preservação da verdade das premissas para a conclusão. Quando uma relação de consequência tarskiana está associada a uma semântica multivalorada esta característica se mantém não em termos da preservação do valor-de-verdade Verdadeiro, mas da preservação de um conjunto de valores-de-verdade ditos designados, que é um subconjunto do conjunto de valores-de-verdade da semântica.

Um exemplo de lógica multivalorada é a Lógica do Paradoxo de Graham Priest. Nesta lógica a semântica possui três valores-de-verdade (Verdadeiro, Falso e Verdadeiro-e-Falso) e destes, dois são designados (Verdadeiro e Verdadeiro-e-Falso). A consequência lógica é definida como a preservação tanto do valor "Verdadeiro" quanto do valor "Verdadeiro-e-Falso" resultando em uma lógica não-clássica paraconsistente. Devemos notar que a Lógica do Paradoxo possui a mesma noção de definição de consequência lógica que a Lógica Clássica, que é a noção tarskiana.

A concepção de multivaloração na lógica moderna foi introduzida pelo lógico polonês Jan Łukasiewicz, que acrescentou o valor-de-verdade Indeterminado aos valores fregeanos tradicionais de Verdade e Falsidade. De acordo com Łukasiewicz, o Indeterminado seria um passo além da dicotomia Aristotélica entre o ser e o não-ser. A noção de multivaloração, no entanto, recebeu duras críticas do também lógico polonês Roman Suszko.

Suszko defendia a tese de que existem "apenas dois valores-de-verdade". De acordo com Suszko, as ideias de Łukasiewicz sobre multivaloração se baseavam em uma confusão entre valores algébricos —aquilo que as sentenças denotam— e valores lógicos —a Verdade e a Falsidade. Sendo assim, os três valores-de-verdade da semântica criada por Łukasiewicz seriam valores algébricos e não valores lógicos.

A denominada Tese de Suszko defende, portanto, a distinção acerca da natureza dos valores-de-verdade. Haveriam dois níveis semânticos para os valores-de-verdade: o nível ontológico relativo à denotação das sentenças e o nível lógico relativo à verdade e à falsidade. De acordo com Suszko, o conjunto de valores-de-verdade, $\mathscr{V}$, representa o conjunto de valores algébricos, em que cada valor denotaria as possíveis referências das sentenças. O conjunto de valores designados, $\mathscr{D}$, e seu complemento, o conjunto de valores não-designados, $\mathscr{U}$, representariam os valores-de-verdade lógicos propriamente ditos (Suszko, 1975a; Suszko, 1977).

A concepção de valor-de-verdade clássica só poderiam ser negada no nível ontológico com o acréscimo de novos valores algébricos, mas permaneceria no nível lógico pela própria noção de relação de consequência tarskiana, cuja definição depende da partição dos valores-de-verdade entre designados e não-designados.

Os valores-de-verdade denotariam, portanto, duas coisas diferentes: os elementos de $\mathscr{V}$ e as partições de $\mathscr{V}$. Desta maneira, adotaremos o termo *lógica multivalorada* para designar uma lógica cuja a cardinalidade do conjunto de valores-de-verdade seja maior do que dois e *lógica multivalente* para designar uma lógica cujo conjunto de valores-de-verdade da semântica tenha mais de uma partição.

Dada a insatisfação com as lógicas multivaloradas, Suszko buscou mostrar formalmente como toda semântica multivalorada, a qual uma relação de consequência tarskiana está associada, pode ser caracterizada de forma bivalente. A chamada Redução de Suszko teve como ponto de partida a descrição bivalorada da lógica de Łukasiewicz em (Suszko, 1975b), mas foi com Malinowski que foi demonstrado que a Redução de Suszko poderia ser aplicada a qualquer lógica multivalorada. Somente mais tarde foi desenvolvido um procedimento que encontra uma semântica bivalente adequada para qualquer semântica multivalorada em (Caleiro et al., 2003)[1].

A *quasi-consequence*, ou *q-consequence*, surge da motivação dada pela ontologia de Łukasiewicz de que haveriam situações indeterminadas. Contestando a Tese de Suszko, Grzegorz Malinowski foi o primeiro a ir além da partição dos valores-de-verdade entre designados e não-designados apresentando uma semântica multivalente associada a uma noção não tarskiana de relação de consequência. Ao criar a *q-consequence*, Malinowski (1990) tripartíciona o conjunto dos valores-de-verdade tomando dois subconjuntos disjuntos: os valores Aceitos e os valores Rejeitados. Como a união dos valores Aceitos com os valores Rejeitados não precisa ser igual ao conjunto dos valores-de-verdade, há margem para que existam valores-de-verdade nem Aceitos e nem Rejeitados.

[1] Sobre a Tese e a Redução de Suszko cf. (Molick, 2015).

A definição de *q-consequence* é dada por: um enunciado é válido se dadas premissas não-Rejeitadas, a conclusão é Aceita. Ou ainda, é incompatível que todas as premissas não sejam Rejeitadas e a conclusão não seja Aceita. Esta noção não é reflexiva, dado que um enunciado de valor nem Aceito e nem Rejeitado não pode ser consequência dele mesmo, pois pela definição é incompatível que a premissa não seja Falsa e a conclusão não seja Verdadeira.

Em (Frankowski, 2004), Szymon Frankowski apresenta uma versão diferente de tripartição dos valores-de-verdade e introduz uma noção de relação de consequência dual à *q-consequence* chamada *plausible-consequence*, ou *p-consequence*. A motivação da *p-consequence* é de ser um raciocínio que lida com conjecturas. Sua semântica associada possui valores Aceitos ou valores Rejeitados, sendo possível haver valores que sejam ao mesmo tempo Aceitos e Rejeitados. A definição de *p-consequence*, é dada por: um argumento é válido se dadas premissas Verdadeiras, a conclusão é não-Falsa. Ou, é incompatível que todas premissas sejam Verdadeiras e a conclusão Falsa.

A *p-consequence* é reflexiva, mas não transitiva. Para mostrar que a transitividade não vale, suponha que as premissas de uma dada consequência são Verdadeiras e estas acarrete somente sentenças cujo valor é Indeterminado. E suponha também que parte destas sentenças sejam premissas de um segundo argumento, e estas acarretem uma consequência Falsa. Como não pode ser o caso que as premissas sejam Verdadeiras e a consequência Falsa, a transitividade não pode ser uma propriedade desta noção de consequência.

Seguindo a ideia de introduzir novas partições ao conjunto de valores-de-verdade, novas semânticas multivalentes foram geradas. E com novas semânticas multivalentes, novas relações de consequência que generalizam a noção de relação de consequência tarskiana são definidas. As questões que deixaremos em aberto para futuros trabalhos é: Como seriam estas novas noções de relações de consequências associada às semânticas multivalentes? O que estas noções teriam em comum com a noção de relação de consequência tarskiana? O que fariam estas noções serem de fato relações de consequência lógicas??

Referências

Beal, J. & Restall, G. *Logical Pluralism*. Oxford University Press, 2006.

Caleiro, C.; Carnielli, W.A.; Coniglio, M.E. & Marcos, J. Suszko's thesis and dyadic semantics. *Preprint available at: <http://wslc.math.ist.utl.pt/ftp/pub/CaleiroC/03-CCCM-dyadic1.pdf>*, 2003.

Dummett, M. *The Logical Basis of Metaphysics*. Harvard University Press, 1991.

Frankowski, S. Formalization of a plausible inference. *Bulletion of the Section of Logic*, v. 33, p. 41–52, 2004.

Malinowski, G. *q*-consequence operation. *Reports on mathematical Logic*, v. 24, n. 1, p. 49–59, 1990.

Marcos, J. Ineffable inconsistencies. In: *Logics of Formal Inconsistency*. Lisboa: Fundação Biblioteca Nacional, 2005. p. 291–300.

Molick, S. *Of madness and many-valuedness: an investigation into Suszko's Thesis*. Dissertação (Mestrado), UFRN, 2015.

Peirce, C. S. *Collected Papers of Charles Sanders Peirce (1931-1958)*. [S.l.]: Harvard University Press, 1974. v. 5.

Rumfitt, I. Knowledge by deduction. *Grazer Philosophische Studien*, v. 77, p. 61–84, 2008.

Shoesmith, D. J. & Smiley, T. J. *Multiple-Conclusion Logic*. Cambridge, MA: Cambridge University Press, 1978.

Suszko, R. Abolition of the fregean axiom. In: *Logic Colloquium*, 1975. p. 169–239.

Suszko, R. Remarks on lukasiewicz's three-valued logic. *Bulletin of the Section of Logic*, v. 4, n. 3, p. 87–90, 1975.

Suszko, R. The fregean axiom and polish mathematical logic in the 1920's. *Studia Logica*, v. 36, p. 373–380, 1977.

Tarski, A. On the concept of logical consequence. In: Corcoran, J. (Ed.). *Logic, Semantics, Metamathematics: papers from 1923 to 1938*. Hackett, 1983 (1936). p. 409–420.

Wittgenstein, L. *Tractatus Logico-Philosophicus (Logisch-philosophische Abhandlung)*. Ogden/Ramsey and Pears/McGuinness English translations, 2015 (1921). Disponível em: <http://people.umass.edu/phil335-klement-2/tlp/tlp.html>.

GALEN STRAWSON, ROBERT HANNA E A METAFÍSICA DA CONSCIÊNCIA

José Sérgio Duarte da Fonseca *

Resumo: Foram submetidas ao exame duas metafísicas da consciência, com características bastante distintas: o *realismo monista* de Galen Strawson e o *pós-fundamentalismo* de Robert Hanna. Apesar de suas diferenças, ambas têm em comum duas premissas, primeiro, que as teorias tradicionais da emergência como superveniência das propriedades experienciais a partir das propriedades físicas são falhas por terem como base a visão fisicalista do mundo; segundo, que o fisicalismo deve ser abandonado em favor de uma visão metafísica do mundo mais apropriada. O realismo monista proposto por G. Strawson sustenta que a emergência só pode ser metafisicamente inteligível a partir de uma visão, de todo modo, pampsiquista do mundo. Hanna, por sua vez, sustenta que, sob bases pós- fundamentalistas, e dispensando o pampsiquismo, é possível explicar a emergência não em termos de superveniência, e sim como emergê ncia dinâmica, i.e, onde propriedades experienciais são fundidas com as propriedades físicas. Mostrei que a crítica de G. Strawson à emergência como superveniência pode ser estendida à emergência dinâmica, a despeito do fato de que a teoria da emergência proposta por Hanna assumir bases bem distintas das teorias tradicionais da superveniência. Como resultado, mostrei que o pós- fundamentalismo de Hanna não é capaz de produzir uma teoria da consciência que inclua a emergência sem que seja necessário incluir também o pampsiquismo para se tornar metafisicamente inteligível.

Palavras-chave: Metafísica. Mundos Possíveis. Consciência. Emergência. Filosofia da mente.

Abstract: Two types of metaphysics of consciousness have been examined, each of them presenting very distinct features: Galen Strawson's realistic monism and Robert Hanna's post-foundationalism. Despite their differences, both positions have in common two tenets: firstly, the assumption that the traditional theories of emergence as supervenience of experimental properties originated from physical properties are not valid since they are grounded in a physicalist vision of the world; secondly, physicalism should be abandoned in favor of a more appropriate metaphysical vision of the world. The realistic monism proposed by G. Strawson claims that the emergence only could be metaphysically grasped if it maintains a panpsychistic vision of the world. Hanna, on his turn, claims that under post-foundationalism, and disposing of panpsychism, it is possible to explain the emergence not in terms of supervenience, but as a dynamic emergence, ie, one in which experimental properties are fused with physical properties. I claimed that G. Strawson's criticism to emergence as supervenience could be extended to dynamic emergence, despite the fact that the theory of emergence suggested by Hanna assumes very distinct bases of the traditional theories of supervenience. As a result, I claimed that Hanna's post-foundationalism is not capable of eliciting a theory of consciousness that includes emergence without having to include panpsychism as well, in order to be metaphysically understandable.

Keywords: Metaphysics. Possible Worlds. Conscience. Philosophy of Mind.

*Professor de Filosofia da Universidade Federal do Piauí e Doutor(2003) em Filosofia pela PUCRio.

Introdução

Uma das maiores dificuldades que qualquer posição não reducionista em filosofia da mente precisa enfrentar é produzir um modelo explicativo de como se dão as relações entre as propriedades da experiência consciente e as propriedades físicas. A estratégia comumente usada para responder a essa dificuldade é mostrar que as propriedades experienciais "surgem" das propriedades físicas, seguindo a intuição academicamente educada que nos diz que a matéria organizada apropriadamente em cérebros ou, talvez mais corretamente, em corpos vivos, "faz surgir" a experiência consciente. Comumente isso se faz a partir da adoção de uma imagem do mundo que é fundamentalmente fisicalista, acrescida da suposição de que a coerência interna de tal imagem manter-se-á intacta mesmo quando se lhe acrescenta a experiência consciente.

Serão examinadas duas teorias da consciência que pressupõem a necessidade de uma crítica ao fisicalismo por ser metafisicamente adequado para a produção de um quadro geral do mundo onde a consciência e o físico possam ser incluídos de forma coerente. Examinarei duas metafísicas da consciência bem como sua pretensão de operar tal reforma. Trata-se de duas tentativas paralelas, sem relação entre si: o Realismo Monista de Galen Strawson (2008) e o Pós-fundamentalismo proposto por Robert Hanna (HANNA e MAIESE, 2009).

Hanna pretende oferecer uma metafísica da consciência a partir da qual seja possível caracterizar o que significa dizer que o físico "faz surgir" o mental, através de uma crítica às teorias da emergência como superveniência e da apresentação de uma nova teria da emergência, a emergência dinâmica das propriedades físicas e mentais em termos de sua fusão em uma corporificação adequada. Esta teoria da emergência, por sua vez, está vinculada a uma teoria mais geral, a teoria da corporificação essencial da consciência.

Em um importante artigo, intitulado "Realistic Monism: Why Physicalism Entails Panpsychism", Galen Strawson (2008)[1] defende, entre outras coisas, a tese de que qualquer teoria da emergência de fenômenos experienciais a partir de fenômenos não experienciais irá fracassar, devido ao fato de que tais teorias não são capazes de explicar "em virtude do que" a emergência ocorre. Sem tal explicação, qualquer teoria da emergência do mental a partir do físico é sempre emergência bruta, ou seja, equivalente a um milagre.

[1] Este artigo, publicado em 1996 no Journal of Consciousness Studies, foi republicado em 2006 em um livro, bem como quinze resenhas críticas do artigo em questão. (FREEMAN, 2006).

No que se segue mostrarei que a crítica de G. Strawson à emergência como superveniência pode ser estendida à emergência dinâmica, a despeito do fato de que a teoria da emergência dinâmica proposta por Hanna assumir bases bem distintas do que as teorias tradicionais. Como resultado, mostrarei que a teoria de Hanna e Maise não é capaz de produzir uma teoria da consciência que inclua a emergência sem incluir também o pampsiquismo para torná-la metafisicamente inteligível.

O realismo monista e a crítica de Galen Strawson à emergência do experiencial a partir do não-experiencial

Strawson parte da tese de que a experiência consciente um fenômeno real e, assim, que nenhum fisicalismo digno deste nome pode deixar de levá-la em conta sem comprometer sua estabilidade teórica. A defesa de tal posição normalmente está vinculada a um tipo de fisicalismo não redutivo. No entanto, Strawson é um crítico severo de qualquer tipo de fisicalismo não redutivo que esteja comprometido com a suposição, para ele insustentável, de que a experiência consciente emerge de fenômenos físicos, essencialmente não experienciais.

A instabilidade teórica do fisicalismo não redutivo se deve ao fato de seus defensores precisam compatibilizar duas teses que, para Strawson, são irreconciliáveis, a saber:

Tese 1:os eventos físicos são essencialmente não experienciais

Tese 2: a experiência consciente é um fenômeno real concreto e, como todo fenômeno real concreto é físico, a experiência consciente é física.

Ao assumir ambas as teses como verdadeiras, o fisicalista não redutivo se vê obrigado a mostrar de que forma os fenômenos experienciais emergem de fenômenos físicos, essencialmente não experienciais.

Strawson considera que a noção de emergência é problemática quando se trata da relação entre o físico e o mental, muito embora ela não o seja quando se trata de determinadas relações entre o físico e o físico, p.ex., o estado líquido da água emergindo das propriedades das moléculas de H2O. Isso se deve ao fato de que a relação de emergência entre propriedades físicas de nível superior a partir de fenômenos físicos de nível mais básico é conceitualmente homogênea e inteligível. No entanto, a emergência de propriedades experienciais, de ordem superior, a partir de fenômenos físicos de nível mais básico é para ele ininteligível, mas não em termos epistêmicos, e sim em termos metafísicos, ou como Strawson prefere dizer, "ininteligível mesmo para Deus".

A dificuldade inerente à emergência dos fenômenos experiências a partir de determinadas propriedades de fenômenos não experienciais, diferentemente da emergência do líquido do não líquido, se deve ao fato de que, se isso é correto, então os fenômenos experienciais são uma mera aparência, o que é absurdo, porque, nas palavras de Strawson, "Eles [os fenômenos experienciais] são podem ser mera aparência, devido ao fato de que toda a aparência depende de sua existência" (STRAWSON, 2008, p. 64).

O fracasso de toda tentativa de explicar os fenômenos experienciais em termos de propriedades não experienciais de nível mais básico fracassaria, devido ao fato de que, se Y emerge de X, então deve haver alguma propriedade em X, e apenas em X, em virtude da qual Y emerge, sendo que tal propriedade é suficiente para Y. A liquidez da água é uma propriedade emergente do comportamento das moléculas de H2O, sujas interações são descritas através de leis apropriadas. Podemos dizer que a liquidez da água é uma propriedade se dá em virtude do comportamento das moléculas de H2O. No caso da emergência do experiencial do não experiencial, ela se daria de forma bruta, já que não há nada que possamos reputar ao não experiencial que nos explicaria em virtude do que o experiencial emergiria dele. Como a emergência é uma relação "em virtude de", tal como Strawson prefere dizer, então não há tal coisa como emergência bruta ou "em virtude de nada", o que equivaleria a um milagre:

> If it really is true that Y is emergent from X then it must be the case that Y is in some sense wholly dependent on X and X alone, so that all features of Y trace intelligibly back to X (where 'intelligible' is a metaphysical rather than an epistemic notion). Emergence can't be brute. It is built into the heart of the notion of emergence that emergence cannot be brute in the sense of there being absolutely no reason in the nature of things why the emerging thing is as it is (so that it is unintelligible even to God). For any feature Y of anything that is correctly considered to be emergent from X, there must be something about X and alone in virtue of which Y emerges, and which is sufficient for Y. (STRAWSON, 2008, p. 65)

Uma defesa possível da emergência do experiencial a partir do não experiencial é dizer que o não experiencial é proto experiencial, ou seja, *intrinsecamente* ajustado a constituir, em certas circunstâncias, fenômenos experienciais. Ocorre que a dificuldade anterior não é contornada com o recurso ao não experiencial, pois, de qualquer forma, dizer que o não experiencial é intrinsecamente ajustado para constituir, em certas circunstâncias, o experiencial é apenas dizer que há algo na natureza do não experiencial *em virtude do que* o não experiencial é assim ajustado. Como não há tal coisa, a emergência do experiencial do proto experiencial é também bruta, e deve ser abandonada.

A solução de Strawson para essa dificuldade está em abandonar a tese 1, a de que os eventos físicos não são essencialmente experienciais, e manter a tese 2, a de que a experiência consciente é um fenômeno real concreto e, como todo fenômeno real concreto é físico, então a experiência consciente é física, abraçando o que chama de realismo monista.

Uma característica do realismo monista de Strawson é que ele pressupõe o pampsiquismo. Se a emergência do intrinsecamente experiencial a partir do intrinsecamente não experiencial é impossível, então o que de fato ocorre quando cérebros, ou melhor, corpos suficientemente complexos estão disponíveis, é a emergência de fenômenos experienciais, ou seja, a consciência como a nossa, a partir de fenômenos físicos que têm consciência *distinta* da nossa. Por consciência como a nossa entendo a consciência de animais humanos e não humanos, acompanhando a definição de Hanna e Maiese (2009) A emergência agora seria metafisicamente inteligível, já que a consciência como a nossa emerge em virtude do caráter essencialmente ajustado das consciências distintas das nossas, ou seja, aquelas dos microconstituintes do cérebro, ou melhor, do corpo. Os fenômenos físicos a partir dos quais os fenômenos experienciais emergem são, eles próprios, experienciais. A consciência como a nossa emergiria da consciência distinta da nossa:

> That is what I believe: experiential phenomena cannot be emergent from wholly non-experiential phenomena. The intuition that drives people to dualism (and eliminativism, and all other crazy attempts at wholesale mental-to-non-mental reduction) is correct in holding that you can't get experiential phenomena from P phenomena, that is, shape-size-mass-charge-etc. phenomena, or, more carefully now — for we can no longer assume that P phenomena as defined really are wholly non-experiential phenomena — from non-experiential features of shape-size-masscharge-etc. phenomena. So if experience like ours (or mouse experience, or sea snail

> experience) emerges from something that is not experience like ours (or mouse experience, or sea snail experience), then that something must already be experiential in some sense or other. It must already be somehow experiential in its essential and fundamental nature, however primitively or strangely or (to us) incomprehensibly; whether or not it is also non-experiential in its essential nature, as conventional physicalism supposes. (STRAWSON, 2008, p. 70.)

Hanna e Maiese partem de uma posição bem distinta da de Strawson, que chamam de pós-fundamentalismo. O pós-fundamentalista afirma que a matéria *pode* estar intrinsecamente ligada à consciência, dadas determinadas circunstâncias, mantendo-se a tese de que a consciência essencialmente distinta do físico. Tais circunstâncias são descritas por Hanna e Maiese em termos de um tipo específico de emergência, onde ocorre a *fusão* das propriedades experienciais com as propriedades físicas, quando da constituição do organismo animal. Resta saber se a emergência dinâmica do pós-fundamentalista recai no mesmo tipo de dificuldades apontadas por Strawson, possibilitando e extensão de sua crítica à tese da emergência do experiencial do proto-experiencial ao pós-fundamentalismo. Em outras palavras, precisamos saber se a emergência dinâmica é um tipo de emergência bruta. Votaremos a isso após uma rápida exposição da teoria da corporificação essencial da consciência.

Pós-fundamentalismo, emergência dinâmica e fusão de propriedades

O pós-fundamentalismo advogado por Hanna e Maiese parte da tese de que a consciência emerge do físico, negando o que eles consideram problemático nas teorias tradicionais da emergência. Os problemas com respeito à emergência surgem, segundo eles, devido à visão fundamentalista do físico e do mental, que precisa ser corrigida em favor ao pós-fundamentalismo[2]. O fundamentalismo se caracteriza pela suposição de que os fenômenos mentais são fundamentalmente mentais e os fenômenos físicos são fundamentalmente físicos e pode ser apresentado da seguinte maneira:

(1) only physical events can cause physical events,

[2] Não vou expor a crítica de Hanna e Maiese às várias teorias da emergência. Cf. HANNA e MAIESE, 2009, pp. 255 ss). Restringir-me-ei à tarefa de descrever a teoria da emergência dinâmica a qual eles chegam depois das críticas às teorias tradicionais da emergência.

(2) a physical event is any real occupant of spacetime that possesses some fundamental physical properties, and

(3) fundamental physical properties necessarily exclude inherent or intrinsic connections with fundamental mental properties. Now to say that something has a fundamental physical property, and thereby necessarily excludes inherent or intrinsic connections with fundamental mental properties, is to say that this thing is fundamentally physical. (HANNA e MAIESE, p. 299)

As várias posições em filosofia da mente podem ser descritas a partir do fundamentalismo: o dualismo, o materialismo eliminativista, o materialismo redutivo (que Hanna e Maiese chamam de fisicalismo) e o materialismo não redutivo, a primeira delas pelo compromisso ontológico com respeito à existência do fundamentalmente físico e do fundamentalmente mental, a segunda limitando tal compromisso apenas com respeito ao fundamentalmente físico sendo as duas últimas obtidas a partir de modalizações específicas com respeito às relações entre o físico e mental, em geral, dispensando o compromisso ontológico com o fundamentalmente mental. Todas essas posições fundamentalistas precisam ser rejeitadas em favor do pós-fundamentalismo. Uma vez caracterizado o fundamentalismo, o pós-fundamentalismo é construído a partir da adoção de (i) e de (ii) e a rejeição de (iii), em favor de:

(iii)* fundamental physical properties do not necessarily exclude any inherent or intrinsic connections with fundamental mental properties, and it is both metaphysically possible and also actually the case that fundamental physical properties include inherent or intrinsic connections with fundamental mental properties, is what we call Post-Fundamentalism. (Idem, p. 300)

Essas conexões "inerentes ou intrínsecas" são definidas com precisão em termos da fusão entre as propriedades fundamentalmente físicas e as fundamentalmente mentais. A fusão de propriedade físico-mental é definida por Hanna e Maiese nos seguintes termos:

(1) Under an embodiment E, an event or physical substance X has some fundamental mental properties M1, M2, M3, etc.

(2) Under the same embodiment E,X also has some non-identical or distinct fundamental physical properties P1, P2, P3, etc.

(3) For every Mi there is a one-to-one correlation with a corresponding Pi.

(4) The members of each 1-1 correlated Mi-Pi pair are necessarily co-extensive.

(5) The members of each 1-1 correlated Mi-Pi pair are not logically necessarily co-extensive.

(6) The members of each 1-1 correlated Mi-Pi pair are mutually inherent or intrinsic structural properties of X.

(7) X is a suitably complex living organism. (HANNA e MAIESE, p 354)

Isso significa que a fusão físico-mental é exemplificada apenas em um mundo constituído por sistemas dinâmicos devido a (5), que diz que as propriedades mentais e propriedades físicas são necessariamente co-extensivas metafisicamente, e não logicamente. Isso elimina de plano a tanto a identidade de propriedades, por exigir a necessidade lógica da co-extensão de propriedades, quanto a superveniência lógica, que requer a relações lógicas de suficiência. (HANNA e MAIESE, p. 328. Para a distinção entre superveniência e fusão, veja também pp. 305-6).

Em outras palavras, a adoção do materialismo pós-fundamentalista nos dá o conforto metafísico necessário para afirmar que a matéria pode, dada uma situação correta, estar essencialmente vinculada, ou seja, fundida ao mental. Tal situação se dá quando a fusão das propriedades mentais e físicas que, como resultado de sua emergência dinâmica da complexidade física e biológica, dá origem a um organismo animal, i.e., essencialmente físico e mental.

Isso, por si só, não parece garantir que tal organismo é metafisicamente um sistema consciente, se adotarmos da perspectiva de terceira pessoa. Podemos seguir Dennett e dizer que podemos, em proveito do entendimento do comportamento de tal organismo, atribuir-lhe intencionalidade, sem qualquer compromisso metafísico com respeito à existência de uma mente consciente. Hanna e Maiese afirmam, no entanto, que a existência da consciência como uma propriedade intrínseca do organismo é metafisicamente necessária, uma necessidade *a priori* fortemente metafísica (*Strong metaphysical a priori necessity*), ou seja, o mundo atual é constituído de tal forma a abrigar consciências essencialmente corporificadas como a nossa, e isso não pode ser conhecido apenas empiricamente, e sim *a priori*:

> Furthermore, because the constraints are universal, intrinsic, and structural, they cannot be known by empirical means alone and thus are a priori. (HANNA e MAIESE, 2009, p. 306)

O pós-fundamentalismo é a metafisica necessária à tese do animalismo corpo- mente, a partir da qual eles propõem seu hilemorfismo neo-aristotélico. O animalista corpo-mente assume que as propriedades físicas e mentais são necessariamente fundidas, e que, conjuntamente, constituem organismos vivos com consciências como a nossa.

A corporificação essencial da consciência é uma das teses centrais da teoria de Hanna e Maiese. Criaturas com consciência como a nossa são completamente corporificadas e de modo metafisicamente necessário.

Mentes desencarnadas são logicamente possíveis, mas, embora concebíveis, não são consciências como as nossas, por isso, estão fora do alcance e do interesse da teoria. A consciência é essencialmente corporificada em dois sentidos, correspondendo a duas teses da teoria de Hanna e Maiese:

> (1) Tese da corporificação essencial da consciência: Necessariamente, a consciência humana é corporificada, i.e., a consciência humana tem uma encarnação neurobiológica em larga escala de todos os seus estados em todos os sistemas vitais e órgãos vitais do animal humano consciente — incluindo o cérebro, o sistema nervoso, o sistema límbico o sistema cardiovascular até os limites da pele, sem ultrapassá-la. (Idem, 50)

A teoria de Hanna e Maiese inclui outra tese, a do animalismo corpo-mente, a partir da qual propõem uma versão renovada do hilemorfismo aristotélico:

> (2) Tese do animalismo corpo-mente: necessariamente, as propriedades físicas e mentais são fundidas, e que, conjuntamente, constituem organismos vivos com consciências como a nossa. Segundo o animalismo, nossas mentes animam nossos corpos neurologicamente complexos guiando propositadamente sua ação. (HANNA e MAIESE, 2009, p. 255)

Os organismos são descritos por Hanna e Maiese como sistemas dinâmicos, assim como o Big Bang, os buracos negros, os sistemas climáticos e os sistemas de tráfico. Uma característica básica dos sistemas dinâmicos é a de que eles são exemplos do que Hanna e Maiese chamam de "singularidades causais naturais". Sistemas dinâmicos são singularidades causais naturais devido ao fato de que o exercício de seus poderes causais se dá a partir de leis estruturais próprias, internas ao sistema, de forma quase independente das leis deterministas e estatísticas que governam as propriedades físicas: "... they naturally create their own future by what they actually do in the present..." (HANNA e MAIESE, p. 262, itálico no original). Consciências como as nossas, essencialmente corporificadas, por serem responsáveis pelos movimentos intencionais do organismo, são um exemplo de singularidades causais naturais[3].

A emergência dinâmica é um tipo de emergência bruta

Há ao menos uma diferença importante entre o pós-fundamentalismo advogado por Hanna e Maiese e o realismo monista de Strawson. Como Strawson, eles também negam a Tese 1, de que os eventos físicos são essencialmente não experienciais, mas vão além do realismo monista, negando também a Tese 2, de que a experiência consciente é um fenômeno real concreto e, como todo fenômeno real concreto é físico, a experiência consciente é física. Por sustentar a Tese 2, o realismo monista de Strawson seria um tipo de fundamentalismo, se visto a partir da perspectiva de Hanna e Maiese. No entanto, há similaridades importantes entre as duas posições com relação à experiência consciente, e tais similaridades serão em proveito do realismo monista.

O argumento de Strawson contra a emergência superveniente do experiencial do não-experiencial está fundado em uma característica essencial da experiência, a de que os fenômenos experienciais, por se esgotarem completamente na aparência, não podem ser mera aparência, o que torna ininteligível qualquer tentativa de explicar a emergência do experiencial do não-experiencial (STRAWSON, 2008). Hanna e Maiese oferecem uma caracterização da consciência bastante particular mas que, apesar disso, não implica na não aceitação da caracterização de Strawson, no que diz respeito ao esgotamento do experiencial na própria aparência.

A posição de Hanna e Maiese com respeito a natureza da consciência é bastante particular pela adoção da Tese da consciência profunda, i.e., necessariamente, se um animal humano consciente possui um algum tipo de estado consciente, então ele possuirá também estados ocor-

[3] De especial interesse é o fato de que, segundo o animalismo corpo-mente, as propriedades mentais não são redutíveis às propriedades físicas, sendo que, as propriedades mentais, por estarem fundidas com as propriedades físicas em um organismo, garantem eficácia causal à mente, já que o organismo exerce seus poderes causais sobre o físico, e tal exercício se dá através de sua ação propositada sobre o meio-ambiente, devido à atuação de suas propriedades mentais. (HANNA e MAIESE, 2009, pp. 256 ss))

rentemente conscientes, mesmo que de forma mínima. Em outras palavras, a tese propõe que todos os estados mentais são conscientes: “consciousness penetrates into every aspect of our mental lives” (HANNA e MAIESE, 2009, p. 28).

Uma importante implicação desta tese é a de que não há algo como o “nível subpessoal”, ou seja, tudo o que importa à mente é pessoal. O que normalmente é considerado como estados mentais não conscientes, ou seja, aqueles que envolvem algum grau de desvio da consciência desperta normal são entendidos, na perspectiva de Hanna e Maiese, como estados conscientes minimamente ocorrentes (ou pré-reflexivos). Dessa forma, todo o processamento de informação é minimamente e ocorrentemente consciente, ou seja, todos os “processos subpessoais” são de fato pessoais, ou seja, acessíveis em primeira pessoa e conscientes. (HANNA e MAIESE, 2009 p. 29). A tese da consciência profunda endossa distinção entre consciência humana reflexiva, racional e a pré-reflexiva, proto-racional, mas não a distinção entre pessoal e subpessoal. Levando-se em conta ambas as teses, a consciência reflexiva, pré-reflexiva e a vida neurobiológica constitui o que somos.

Uma das mais importantes contribuições de Daniel Dennett à filosofia da mente são seus argumentos em prol da eliminação dos qualia, relegando a experiência subjetiva ao mundo virtual (heterofenomenológico), que corresponde ao nível pessoal de descrição de um agente humano, enquanto o que importa de fato entender estaria no nível sub-pessoal: no nível pessoal, o que se tem é um agente cognitivo como uma totalidade; no nível sub-pessoal, o que se tem são as operações do sistema cognitivo interno do agente (DENNETT, 1991). Trata-se de uma distinção aceita de forma quase unanime na neurociência. De fato, o enfoque epistemológico e metodológico das ciências cognitivas está assentado no nível sub-pessoal.

Essa distinção já clássica entre os níveis pessoal e sub-pessoal é posta em questão por Hanna e Maiese (2009), com a proposta de uma teoria da consciência essencialmente corporificada, quer dizer, corporificada não apenas no sistema nervoso, mas no organismo humano como um todo. Hanna e Maiese adotam a postura “qualiófoba” de Dennett, mas, em contraste, Hanna e Maiese também defendem a realidade da experiência consciente: “So we are at once qualia eliminativists and also freaks about consciousnesslo” (HANNA e MAIESE, 2009, pp. 76). Com respeito à experiência consciente, eles supõem que elas são:

(i) intrinsic and also structural properties, i.e., necessary, internal, relational properties that are inherently bound up with the spatiotemporal neurobiological dynamics of our living organismic bodies,

(ii) effable, i.e., communicable to another essentially embodied subject who is suitably egocentrically positioned in orientable space and thermodynamically irreversible time, even if not conceptually describable to that subject,

(iii) shareable, at the very least, by means of empathic mirroring of intentional body movements — i.e., as movement-types, although not as tokens of those movement-types,

(iv) directly apprehensible, i.e., available without further cognitive mediation to either pre-reflectively conscious sensorimotor subjectivity or self-conscious, self-reflective introspective subjectivity, and

(v) fallible, i.e., open to introspective misinterpretation features of all conscious states like ours. (HANNA e MAIESE, 2009, pp. 76-7)

O que é relevante aqui é o fato de que, mesmo adotando uma postura qualiófoba com respeito aos qualia, Hanna e Maiese, ao adorem a tese da consciência profunda, não põem nenhum obstáculo ao que mais importa para Strawson, i.e., já que, segundo eles, não há o sub-pessoal, a partir do qual poder-se-ia sustentar uma revisão da experiência em termos externos a ela, então o experiencial se esgota na aparência.

A tese da emergência dinâmica advoga algo bem próximo da estratégia, criticada por Strawson, de defender a possibilidade do mental a partir das propriedades proto-experienciais do físico. Na terminologia de Strawson, o não experiencial é proto-experiencial se é intrinsecamente ajustado para constituir, em certas circunstâncias, fenômenos experienciais. O mundo constituído de sistemas dinâmicos de Hanna e Maiese seria proto-experiencial do ponto de vista de Strawson; por outro lado, do ponto de vista pós-fundamentalista de Hanna e Maiese, a crítica de Strawson à emergência do mental a partir do físico é restrita à emergência como superveniência, e assim, subalterna à visão mecânica e estratificada do mundo.

A crítica de Strawson à emergência como superveniência pode ser estendida à emergência dinâmica de Hanna e Maiese? A emergência dinâmica de propriedades físico-mentais se dá através da fusão de propriedades físicas e de propriedades mentais em uma corporificação específica, constituindo um novo tipo de sistema dinâmico, o organismo vivo animal. Mas a fusão de propriedades físicos-mentais que resulta da emergência dinâmica *se dá em virtude do que*?

Hanna e Maiese nos dão alguns exemplos de emergência dinâmica, sendo o entrelaçamento quântico um de seus exemplos favoritos: "Em sistemas quânticos entrelaçados, o novo composto resultante determina os constituintes originais (partículas), e não ao contrário, como a superveniência mereológica poderia sugerir" (HANNA e MAIESE, p.365). O que está entrelaçado, ou seja, fundido, em tais sistemas quânticos são determinadas propriedades quânticas das partículas constituintes, propriedades que existiam em separado antes da interação que resultou no entrelaçamento ou fusão de tais propriedades. O entrelaçamento quântico ocorreu em virtude das propriedades preexistentes das partículas antes da interação, em função das leis da mecânica quântica. Este é um exemplo de emergência dinâmica bem-comportada, não bruta, pois fica claro em virtude do que a fusão de propriedades (entrelaçamento quântico) emerge.

E quanto à emergência dinâmica das propriedades físico-mentais no organismo animal? Hanna e Maiese entendem que são as propriedades mentais que emergem dinamicamente, dada uma dada corporificação adequada. Se perguntarmos em virtude do que as propriedades mentais emergem das propriedades físicas, teríamos de responder que as propriedades mentais emergem dinamicamente em uma determinada corporificação, dadas as circunstâncias adequadas.

Hanna e Maiese defendem a tese de que a fusão entre o físico e o mental no organismo animal deve ser entendido de forma *a priori*: o mundo efetivo é constituído de tal fora que as propriedades físicas e as propriedades mentais podem se fundir, dadas as condições adequadas. No entanto, isso nada nos diz sobre a inteligibilidade metafísica da emergência dinâmica: as propriedades experienciais e não experiências são fundidas a partir do que? O entrelaçamento quântico é plenamente inteligível: dadas duas partículas com determinadas propriedades, após a interação adequada, suas propriedades preexistentes se tornam entrelaçadas, ou fundidas, no vocabulário de Hanna e Maiese. Para ser inteligível, no sentido metafísico de Strawson, a fusão das propriedades físicas e mentais quando da constituição do organismo animal deveria ocorrer entre propriedades físicas e mentais preexistentes, o que implicaria algum tipo de visão pampsiquista, à qual eles veemente rechaçam (HANNA e MAIESE, 2009) Assim sendo, a emergência dinâmica preconizada por Hanna e Maise é de fato mais um exemplo de emergência bruta, pois não nos diz, de forma alguma, em virtude do que as propriedades mentais emergem das propriedades físicas.

Conclusão

Meu objetivo foi mostrar que a crítica de G. Strawson à emergência como superveniência pode ser estendida à emergência dinâmica de Hanna e Maiese. Segundo os últimos, a emergência dinâmica de propriedades físico-mentais se dá através da fusão de propriedades físicas e de propriedades mentais em uma corporificação específica, constituindo um novo tipo de sistema dinâmico, o organismo vivo animal. Se perguntarmos em virtude do que as propriedades mentais emergem das propriedades físicas, teríamos de responder, se quisermos seguir Hanna e Maiese, que as propriedades mentais emergem dinamicamente em uma determinada corporificação, dadas as circunstâncias adequadas. No entanto, já que as propriedades mentais são propriedades novas e inauditas, então tais circunstâncias precisam ser fundamentalmente físicas e fundamentalmente mentais, fazendo com que a posição pós-fundamentalista de Hanna e Maiese se colapse no fundamentalismo que eles rejeitam, já que não é possível defender, no interior da teoria, a inteligibilidade metafísica exigida por Strawson no que diz respeito às teorias da emergência do mental a partir do físico. A fusão de propriedade físico-mental que resulta da emergência dinâmica se dá em virtude do que? A teoria de Hanna e Maiese não tem uma resposta adequada à esta pergunta, o que torna a emergência dinâmica mais um tipo de emergência bruta.

Referências

DENNETT, Daniel. *Consciousness Explainedm*. Boston: Little, Brown, & Co., 1991.

HANNA, Robert; MAIESE, Michelle. *Embodied Minds in Action*. Oxford: Oxford University Press, 2009.

STRAWSON, Galen. Realistic Monism: Why Physicalism Entails Panpsichism? In: *Real Materialism and Other Essays*. Oxford: Oxford University Press, 2008.

FREEMAN, Anthony, (ed.). *Consciousness and Its Place in the Nature*. Does Physicalism Entail Panpsychism? Imprint Academic, 2006.

UMA CONTRIBUIÇÃO AO TEMA DA CONSCIÊNCIA CORPORIFICADA E A LÓGICA DO CORPO

Maria Cristina de Távora Sparano *

Resumo: A partir da posição apresentada por (HANNA; MAIESE, 2009), no seu livro *Embodied Minds in Action*, e por (MAIESE, 2011), em *Embodiment, Emotion and Cognition*, e, ainda, das contribuições do livro de Evan Thompson (2007), *Mind in Life: Biology, Phenomenology and Sciences of the Mind*, pretendemos, com os autores, deslumbrar essa nova possibilidade para a relação mente-corpo, salvaguardadas as diferenças básicas próprias de um dualismo que tem como base o modelo cartesiano. A proposta dos autores que pretendemos analisar tem como objetivo situar agentes conscientes em um patamar cognitivo mais abrangente situado no corpo dos indivíduos. Para os autores, criaturas com consciência e intencionalidade são essencialmente corporificadas, as manifestações dessa corporificação essencial é uma experiência subjetiva de *desire-based emotion*. Para os autores, essa posição, diferentemente daquela do enunciado cartesiano "Eu penso, logo sou", diz: *I desire, therefore I am.* Para (HANNA; MAIESE, 2009), contrariamente ao cogito cartesiano que atribui consciência somente a animais racionais, essa consciência é uma capacidade protorracional: animais com mentes (como crianças pequenas ou portadores de doenças que afetam as capacidades cognitivas), sendo suficiente a capacidade de querer desejar, o que para os autores e especialmente para mim, é o princípio da capacidade de pensar pois para pensar, é preciso querer pensar.

Palavras-chave: Corpo. Consciência. Desejo. Emoção..

Abstract: From the position presented by (HANNA; MAIESE, 2009), in the book *Embodied Minds in Action*; (MAIESE, 2011), in *Embodiment, Emotion and Cognition*; and the contributions from the book of (THOMPSON, 2007), *Mind in Life: Biology, Phenomenology and Sciences of the Mind*, we intend to dazzle a new possibility for the relationship mind-body, safeguarded the basic differences of a specific dualism, which are based on the Cartesian model. The authors' proposal that we intend to analyze aims at settling conscious agents in a more comprehensive cognitive level located in individual's body. For the authors, creatures with consciousness and intentionality are essentially embodied, the manifestations of this essential embodiment is a subjective experience of desire-based emotion. Hanna and Maiese also state that such position, unlike that of the Cartesian statement: I think, therefore I am, goes as: I desire, therefore I am. For Hanna and Maiese, unlike the Cartesian *cogito* consciousness that attaches itself only to rational animals, awareness is a protorational capability: animals with minds (as well as small children or patients with diseases affecting cognitive abilities) bear sufficiently the capacity to want or desire. For the authors and me, that is the same principle of the capacity to think, because in order to think I must want to think.

Keywords: Body. Consciousness. Desire. Emotion..

*Possui graduação em Filosofia pela Universidade Federal do Paraná (1972), mestrado em Filosofia - Université de Paris I - Pantheon Sorbonne (1980), Doutorado Sanduiche em Filosofia- Université de Montréal (1994) e doutorado em Filosofia da Linguagem pela Pontifícia Universidade Católica do Rio Grande do Sul (2002). Pós-doutorado em Filosofia - University of Geneva (2013- 2014). Professora de Filosofia (UFPI).

Introdução

Na via traçada pela dicotomia cartesiana entre pensamento e extensão, o cientificismo purificou o corpo, reduzindo-o a uma máquina que pode ser completamente mapeada, ter seu funcionamento previsto e passível de ser programado, desconhecendo que o corpo está marcado pela **consciência** e por seu **animalismo**.

O objetivo deste trabalho é mostrar o lugar do corpo e suas implicações filosóficas para a cognição em um modelo alternativo de consciência que se situa além do modelo cartesiano clássico de consciência. Com uma possível hipótese para um sujeito pós-cartesiano, mostrarei a natureza da consciência corporificada e suas propriedades; e como esta pode oferecer respostas para a relação mente-corpo e sua ação no mundo. Indagarei sobre a importância da contribuição do modelo proposto para a filosofia da mente, ao passo em que analisarei suas teses e a estrutura disposicional dos estados mentais, como crenças, desejos e emoções baseadas em desejos. O movimento dessa proposta vai de “Eu penso, logo sou” para “Eu desejo, logo sou”, evocando sua importância tanto em relação à sua descrição e à sua função, como à sua atuação no mundo. A diferença entre materialismo cartesiano e corporificação essencial da consciência não se limita ao cérebro ou à materia que situa apenas uma parte do cérebro como consciente e descarta seu animalismo natural que é essencialmente físico, mas também mental, e que, seguindo a direção dos autores (HANNA; MAIESE, 2009), engloba o corpo como um todo e seus desejos.

A relação entre fenômenos físicos e mentais, assim como a indagação sobre a natureza dos corpos e da mente são aspectos dessa relação que incitam novas questões e que mostram a atualidade filosófica das novas pesquisas, principalmente aquelas do âmbito das ciências neurofisiológicas e neurocognitivas e psicológicas.

A corporificação essencial da consciência

Para a pergunta “o que são criaturas com corpos, como nós?”, a resposta é: criaturas neurologicamente corporificadas com um sentido de consciência enraizado no mundo, orientadas de forma dinâmica e organizadas de forma irreversível por estruturas espaciais e temporais, costumeiramente chamadas de pessoas, fornecendo uma hipótese para um sujeito pós-cartesiano vivo e material (HANNA; MAIESE, 2009).

Em decorrência da teoria proposta por (HANNA; MAIESE, 2009) e para nosso uso, essas criaturas são pessoas; e uma pessoa é um objeto físico material e complexo, além disso, "[...] com uma consciência como a nossa essencialmente corporificada." A "Corporificação Essencial da Consciência" não se limita ao cérebro ou ao sistema nervoso central, englobando "nosso corpo como um todo"(HANNA; MAIESE, 2009)[p. 102]. Mas para a pergunta "o que é o corpo de uma pessoa?", a teoria proposta irá dizer que: corpos como os nossos realmente parecem funcionar de um modo mais complexo e curioso do que quaisquer outros objetos materiais conhecidos, sejam eles naturais ou artificiais. Contudo, nossos corpos são apresentados como entidades materiais que podem ser explicadas em termos puramente físicos que se definem por uma proposta essencialmente fisicalista.

Mas podemos oferecer uma visão completa de uma pessoa que dependa totalmente de explicações físicas?

Para muitos, não podemos apresentar uma resposta adequada puramente física de pessoas. Pessoas percebem, têm sentimentos, pensam e agem intencionalmente. Objetos puramente físicos não podem assim agir. Então, qual a diferença entre pessoas e objetos físicos?

Hanna e Maiese

A partir da posição apresentada por (HANNA; MAIESE, 2009), no seu livro *Embodied Minds in Action*, e por (MAIESE, 2011), em *Embodiment, Emotion and Cognition*, e, ainda, as contribuições do livro de (THOMPSON, 2007), em *Mind in Life: Biology, Phenomenology and Sciences of the Mind*, pretendemos, com os autores, deslumbrar essa nova possibilidade para a relação mente-corpo, salvaguardadas as diferenças básicas próprias de um dualismo que tem como base o modelo cartesiano.

A proposta dos autores que pretendemos analisar tem como objetivo situar agentes conscientes em um patamar cognitivo mais abrangente situado no corpo dos indivíduos. Para os autores, criaturas com consciência e intencionalidade são essencialmente corporificadas, as manifestações dessa corporificação essencial é uma experiência subjetiva de *desire-based emotion*. Para os autores, essa posição, diferentemente daquela do enunciado de (**??**)"Eu penso, logo sou", diz: *I desire, therefore I am*. Para (HANNA; MAIESE, 2009), contrariamente ao cogito cartesiano que atribui consciência somente a animais racionais, essa consciência é uma capacidade protorracional de animais com mentes. Chimpanzés, crianças pequenas ou portadores de doenças que afetam as capacidades cognitivas são animais com mentes, sendo suficiente a capacidade de querer ou desejar, o que para os autores é o princípio da capacidade de pensar e, especialmente para (MAIESE, 2011), para pensar, é preciso querer pensar.

A teoria desenvolvida pelos autores apresenta um aspecto fundamental: o animalismo corpo-mente. A hipótese apresentada tem como base filosófica o hylemorfismo aristotélico em uma versão atualizada para animais com mentes. A recusa de um materialismo simplificado para a consciência e de um mentalismo estrito em primeira pessoa faz da proposta uma integração mente-corpo, que os autores chamam de realização autopoiética.

A descrição de criaturas com consciência e intencionalidade essencialmente corporificadas e a manifestação dessa corporificação será examinada como experiência subjetiva. Pode parecer, à primeira vista, que o caráter dessa descrição seja apenas funcional, mas os autores afirmam que há condições suficientes para demonstrar que corpos integrados à natureza e capazes de agir são consciências **essencialmente** corporificadas.

Ao contrário de organismos vivos, são corpos conscientes e dotados de intenção na consecução de uma ação. Diferentemente da neurociência, cujas teorias localizam a consciência no cérebro, a filosofia da mente de Hanna e Maiese localiza-a no agente, sem esquecer suas bases biológicas.

Embora, em princípio, possamos afirmar que ter uma matriz para a consciência nessas bases, isto é, corporificada, não leve a um fechamento da atividade racional, também não considera o elemento racional consciente na forma convencional dada pela tradição, mas inclui no estudo da consciência estados não conscientes e intencionais.

Hanna e Maiese: A consciência é essencialmente corporificada

A proposta de uma teoria da consciência essencialmente incorporada é vinculada a duas teses da teoria de (HANNA; MAIESE, 2009):

a) tese da corporificação essencial da consciência: necessariamente, a consciência humana é corporificada, isto é, a consciência humana tem uma encarnação neurobiológica em larga escala de todos os seus estados em todos os sistemas vitais e órgãos vitais do animal humano consciente - incluindo o cérebro, o sistema nervoso, o sistema límbico, o sistema cardiovascular até os limites da pele e com uma extensão expandida aos domínios do ambiente;

b) b) tese da consciência profunda: necessariamente, se um animal humano consciente possui algum tipo de estado consciente, então ele possuirá também ocorrências conscientes, mesmo que de forma mínima. Em outras palavras, a tese propõe que todos os estados mentais são conscientes: "[...] consciousness penetrates into every aspec to four mental lives" (HANNA; MAIESE, 2009)[p.29]. Não há pensamentos subconscientes, todos os estados e processos mentais pressupõem subjetividade sensório-motora e esse é um modo primitivo de consciência corporal que fornece uma perspectiva pessoal tanto subjetiva como objetiva.

Os autores buscam, ainda, explicitar uma *neurophenomenological analysis of consciousness*, que tem sua fonte em estudos apresentados anteriormente por (HANNA; THOMPSON, 2003). Com a noção de *neurophenomenological analysis* da consciência, queremos explicitar os seguintes pontos do projeto a seguir.

> (i) descrever consciência, inclusive seus vários atos mentais, conteúdo e alvo tal como aparecem em primeira pessoa,
> (ii) estabelecer algumas proposições necessárias e a priori acerca de mentes conscientes com base naquelas descrições em primeira pessoa;
> **(iii) tornar essas proposições tão coerentes quanto possível com a evidência empírica das neurociências cognitivas inclusive a psicologia cognitiva, neurologia médica, neurofisiologia e neurobiologia**[1] (HANNA; MAIESE, 2009)[p. 21] tradução nossa, grifo nosso).

Animalismo corpo-mente

O animalismo corpo-mente defendido por Hanna e Maiese propicia uma versão renovada do funcionalismo, o assim chamado funcionalismo do corpo vivente (*Living Body Functionalism*) e seu correlato, a "corporificabilidade múltipla" (*multiple embodiability*), concebidos a partir da rejeição das versões recebidas do funcionalismo, tidas como reducionistas (HANNA; MAIESE, 2009)[p. 352].

[1] (i) to describe consciousness including its various mental acts, contents, and targets as it appears to the first person,
(ii) to frame some necessary a priori claims about conscious minds on the basis of those first person descriptions,
(iii) to make these claims cohere as closely as possible with empirical evidence from the cognitive neurosciences including cognitive psychology, medical neurology, neurophysiology, and neurobiology. (HANNA; MAIESE, 2009)[p. 21].

O funcionalismo do corpo vivente é por eles definido da seguinte forma: se para um dado sistema dinâmico, as propriedades funcionais determinarem exatamente a mesma eficácia causal de um organismo vivente, então tal sistema dinâmico é um organismo vivente. "Ou, dito de outro modo, [e literalmente] necessariamente, qualquer coisa do mesmo tipo de sistema dinâmico como um corpo animal orgânico é um corpo orgânico animal. Chamemos a esta tese de Funcionalismo do Corpo Vivo"[2] (HANNA; MAIESE, 2009)[p. 343], tradução nossa.

Essa tese, uma versão renovada do hilemorfismo aristotélico, sob esse aspecto, sugere uma oposição ao dualismo cartesiano. As propriedades físicas e mentais são uma fusão, são também sistemas dinâmicos, forças dinâmicas e causais próprias ao organismo vivo, forças que animam um corpo neurologicamente complexo que se move, provocando ações. De especial interesse é o fato de que, segundo o **animalismo corpo-mente**, as propriedades mentais não são redutíveis às propriedades físicas, sua ação conjunta atua sobre o meio ambiente.

Hanna e Maiese defendem como posição filosófica uma metafísica dos **sistemas dinâmicos**, estreitamente vinculada com a visão pós-fundamentalista do mundo, explicitada na tese da emergência dinâmica, segundo a qual o mundo não é nem fundamentalmente físico nem fundamentalmente mental, mas uma totalidade de forças dinâmico-causais, organicamente arranjadas.

A adoção do materialismo pós-fundamentalista nos dá o conforto necessário para afirmar que a matéria pode, dada uma situação correta, estar essencialmente vinculada, ou seja, fundida ao mental. Tal situação se dá quando as propriedades mentais emergem dinamicamente da complexidade física e biológica, constituindo um organismo essencialmente físico e mental.

Emoções baseadas em desejos (*Desire based emotions*)

Dando continuidade a essa tese, (MAIESE, 2011), em outro livro, *Embodiment, Emotion, and Cognition*, explora esses sistemas dinâmicos e os adiciona a um sujeito corporificado. Seu objeto é essencialmente um sujeito corporificado tendo como fundamento suas raízes biológicas e naturais. De acordo com outros autores, como Varela, (VARELA; THOMPSON; RORSCH, 1991) e (THOMPSON, 2007), experiências pessoais partem de um ponto de vista centrado. Esse sujeito não é um sujeito conceitual ou uma experiência autorreflexiva, apenas neurobio-

[2] "Or, in other words, necessarily anything that is the same kind of dynamic system as a living organismic animal body, is a living organismic animal body. Let us call this thesis Living Body Functionalism" (HANNA; MAIESE, 2009)[p. 343]

logicamente centrado. No entanto, o sentido do sujeito e da consciência, por definição, é imanente e reflexivo e essa corporificação essencial terá necessariamente um *conatus sensorimotor; afectus subjetivo* na definição de um sujeito substancialmente corporificado, uma consciência corporificada. Esse modelo de fusão considera as conexões entre o desejo, a emoção, a ação e a experiência corporal. Para Maiese (2011), há, portanto, uma reflexividade imanente, intransitiva, não intrínseca, mas relacional. Um dos pontos a ser explorado é a relação espaço-tempo, em que a espacialidade e o sentido do tempo têm estrutura e propriedades intrínsecas (posição do corpo no espaço e percepção do tempo) num sistema específico (diacrônica e sincrônica). Maiese, para caracterizar a estrutura da consciência, mantém o modelo de bidimensional de Searle como unidade horizontal e vertical, diacrônica e sincrônica. Essas propriedades intrínsecas e relacionais são acidentalmente externas e necessariamente internas, mas globalmente orientadas no espaço e no mundo, são assimétricas e irreversivelmente tomadas no tempo.

Em (HANNA; MAIESE, 2009), são chamadas propriedades estruturais intrínsecas. Os benefícios dessas estruturas espaço-temporais é que elas mantêm relações com o meio ambiente. Portanto, o ponto de origem da consciência é um corpo com uma continuidade de experiências perceptivas numa dimensão de engajamento no mundo.

Para a questão recorrente, - O que são criaturas com corpos como os nossos? -, a resposta da referida autora: são criaturas neurologicamente corporificadas com um sentido primitivo de *self* globalmente enraizado, dinamicamente orientado e irreversivelmente organizado por estruturas espaciais e temporais.

Nosso corpo é um corpo vivido (*lived body*) situado em um mundo onde o sentido, o significado e possibilidades de percepção e ação são fundamentalmente uma questão de desejo baseado em emoções (*desire based emotion*).

O sentido do self: de "eu penso, logo sou" para "eu desejo, logo sou"

Para muitos filósosfos, como Kant, Nagel e Searle, a consciência humana é necessariamente unificada e, ainda mais, essa unidade é o fundamentto da identidade pessoal e de um sentido único e individual do *self*. Alguém que realmente não se preocupasse com sua vida, suas experiências ou atividades em um determinado momento não somente perderia um coerente sentido de si no momento, mas também completamente o de uma vida consciente como a nossa (MAIESE, 2011).

A diferença entre criaturas dotadas de uma consciência cartesiana das que transcenderam essa perspectiva caracteriza-se na medida em que estas são dotadas de um desejo de primeira ordem de dar auto expressão corporal espontânea a cada emoção. Essas emoções são baseadas em desejos, vontades, inclinações, gostos, humores, paixões e sentimentos de pessoas que desejam coisas e que querem relacionar-se com outras, interagir socialmente. Seres com esse *conatus afetivo* são uma constante fonte de energia, de movimento corporal intencional compreendido como apetite, ímpeto, pulsão.

"A ubiquidade do 'eu' baseado em desejos e emoções reflete o fato de que tudo que aparece como algo, é algo que confere sentido [...] e algo pelo qual somos tocados, afetados, estimulados, surpresos e em algum nível violados" [WANDESFELS apud MAIESE, 2011, p. 110, tradução nossa]. Somos provocados por eventos e objetos de toda natureza e, para que tenham relevância, devem ser eventos relacionados ao sujeito, ao "eu". Os modos de consciência corporal que incluem ambos são percepção somática, movimentos enraizados na experiência do corpo de um indivíduo como centro de desejo e de importância.

Esse sentido básico de desejo baseados em emoções (*desire based emotions*) aparece na infância e visa suprir, em um primeiro momento, necessidades básicas e leva a experiências de um "eu" e de um agir próprio. Da mesma forma, na idade adulta, essas experiências tornam-se mais sofisticadas, mas se encontram sempre engajadas numa experiência estimulante para o indivíduo. Mesmo em movimentos involuntários, autores como Sheets-Johnston (apud (MAIESE, 2011)) sustentam que surgem de uma experiência afetiva que brotou de desejos e de emoções.

São desejos pré-reflexivos que, espontaneamente, movem seu corpo, reagindo ou expressando modos de como o indivíduo quer o mundo.[3] Auto expressão corporal significa que mesmo um indivíduo paralisado tem existência corporal, tem noção do movimento corporal e se importa com sua condição, e isso é relevante para a auto expressão. Desejos baseados em emoções são importantes não apenas para o movimento do corpo, mas também quando o movimento falha ou quando por qualquer razão somos incapazes de nos mover.

> Entender o corpo como centro de desejo e de importância também nos permite compreender porque o indivíduo experiência seu próprio corpo como ambos, sujeito e objeto. Como sujeito, posso me mover de modo a conseguir o que desejo e ambas minhas percepções e interpretações são modeladas pelo que desejo, val-

[3] Sintomas que se expressam no corpo do paciente.

> orizo e considero significante. Mas, por outro lado, objetos do mundo lá fora me impactam e me afetam de forma que vão de encontro ou se opõem a meus desejos (MAIESE, 2011)[p. 111], tradução nossa.

O corpo é um corpo vivo e vivido, cujos significados e cujas possibilidades de percepção da ação são fundamentalmente uma questão de emoção e desejos. O sentido do *self*, do "eu", de "mim" e das experiências subjetivas numa perspectiva egocêntrica fundada em desejo baseados em emoções, faz-se através da experiência corporal e tem um fundamento metafísico para situações limites do corpo até então compreendidas e tratadas fenomenológica e clinicamente a partir de seus sintomas; mesmo em situações-limite, como é o caso da agnosognosia[4], quando pacientes declaram que seus corpos e movimentos estão normais, embora paralisados. O importante é que mesmo apesar da falta de percepção em relação ao estado de seus corpos os pacientes nunca declaram ser desprovidos de corpos e retêm algum sentido dos limites corporais (CARRUTHERS, 2008). O paciente se faz doença com seu corpo e seu "eu" está ligado à percepção ou não de seu corpo.

O autor citado sustenta que mesmo representações desconexas do corpo têm um sentido consciente de personificação; e afirma a compatibilidade dessas sensações desconexas com a noção de sintonia corporal de contexto?, o que, para (MAIESE, 2011), é chamado de enquadramento afetivo, que não se apoia na percepção atual do corpo nem numa consciência proprioceptiva, mas em sentimentos personificados e em desejos de enquadramento afetivo (*affective framing*).

Mesmo que os estados corporais, os sentimentos, os pensamentos e os movimentos estejam num fluxo contínuo e, muitas vezes, fenomenalmente desarticulados, eles têm um ponto em comum: a vivacidade da mente essencialmente corporificada que existe como entidade desejosa durante o passar do tempo, envelhece e eventualmente morre.

A noção de mente corporificada e, consequentemente, personificada, cuja raiz é desejo baseado em emoções, captura tanto a espacialidade como a temporalidade do "eu", sua inserção no espaço e no tempo. Assim, podemos concluir que a extensão *corpo* - diferentemente do conceito cartesiano - mais desejo e emoções de pessoas, muitas vezes diagnosticadas como depressivas ou que sofrem de despersonalização, como marcas de caráter, que não se importam com sua vida e suas experiências vividas, são pessoas que mesmo assim experienciam fenômenos

[4] Agnosognosia *refere-se à falta de consciência, deficits sensoriais, motores e alterações cognitivas que ocorrem como consequência direta de lesão cerebral adquirida* (acidente vascular cerebral, trauma cranioencefálico, infecções cerebrais) e doenças neurodegenerativas (demência).

tanto corporais como mentais. Mesmo alguém que, no limite de suas forças psíquicas, sofrendo de transtornos mentais não se importasse com sua vida, com suas experiências, sofrendo uma total falta de coerência do sentido de si e de sua vida nesses momentos, só não seria um sujeito e uma pessoa se lhe faltasse vida consciente como a nossa.

(MAIESE, 2011) descreve esse processo essencialmente corporificado de um ponto de vista mental especial, nomeando-o reflexividade imanente.

> Acredito que isto é uma propriedade estrutural intrínseca de criaturas conscientes, vivas como nós e que é um resultado natural de nossa animada dinâmica neurobiológica. Então o *self* por si mesmo é nada mais, nada menos que um dinâmico, disposto, vivo e essencialmente processo corporificado - com efeito, uma forma viva ou uma forma de vida ((MAIESE, 2011)[p. 111], tradução nossa).

Os impulsos direcionados à auto estima, que (VARELA; THOMPSON; RORSCH, 1991)[p. 62] descrevem como "instintivos, automáticos, difusos e poderosos", são uma expressão de nossos "desejos baseados em emoções", que, por sua vez, são uma expressão de nossa dinâmica neurobiológica. Do mesmo modo, a tendência de nos preocuparmos com o eu, com o *self*, e de nos deixarmos levar pelo que os autores descrevem como "agarrar-se ao ego", está enraizada em nossa dinâmica neurobiológica.

Maiese concorda com essa ideia, ao dizer que agarrar-se ao ego pode, de fato, levar a considerável sofrimento; e acredita que esse tipo de auto estima está no centro daquilo que significa ser um animal humano que se esforça para estar vivo e viver bem.

Conclusão

Um sujeito consciente, ao ter sentimentos em relação às coisas que o rodeiam e ao importar-se ou não com coisas e objetos, ao sentir sua presença, inerentemente também **se importa consigo mesmo** e com seus próprios estados mentais. É a experiência contínua do indivíduo ao importar-se com seu ponto de vista corporificado particular que dá a criaturas como nós um sentido de auto continuidade e coerência, pois, justamente como (VARELA; THOMPSON; RORSCH, 1991) observam, um indivíduo experimenta raiva ou se assusta quando é ameaçado; um indivíduo torna-se ganancioso ou anseia por auto aprimoramento ou, ainda, pode tornar-se entediado quando a situação com a qual está lidando parece irrelevante. Essa experiência contínua de importar-se, preocupar-se, que é a base para todas as percepções, sensações corporais, sentimentos e motivações de um indivíduo, é biologicamente fundada e não pode ser extinta pela perda de partes específicas do corpo ou se numa postura dualista mente-corpo as desconsiderarmos como parte integrante da dinâmica corporal.

A extinção de desejos, emoções, razões para agir só se dá pela completa destruição da dinâmica neurobiológica do indivíduo, ou seja, a morte biológica e psíquica. É a plasticidade de um corpo consciente que dá o sentido de seu corpo e corpos entre outros corpos apresentada em caráter extremo por algumas patologias apresentadas por Maiese (agnosognosia, autismo, esquizofrenia)[5] são do mais vivo interesse para nós nessa pesquisa, pois diríamos que a descrição fenomenológica de mundo circundante (*Umwelt*) é fundamental para a análise desse contexto, mas por si só não é suficiente, assim como as respostas neurofisiológicas ou neurológicas que nos indicam uma postura fisicalista.

Corpos dotados de uma consciência como a nossa numa dinâmica de desejos e emoções são corpos vivos, situados num mundo cujo significado, importância e possibilidades de percepção e ação são fundamentalmente uma questão de desejos baseados em emoções e suportam relações de significado de sujeito e sujeitos no mundo.

Essa teoria é para nossa reflexão uma teoria provocativa e desafiadora porque, além de articular as emoções com a mente, mostra como estas só se dão e existem num corpo. A teoria da mente corporificada e a ligação da mente com o corpo num processo de vida e atividade do corpo suporta a ideia de que o corpo na sua totalidade e abrangência e, nao é apenas o cérebro que modela e estrutura a mente numa postura reducionista, estritamente fisicalista. Querer ou desejar entrelaçam as partes e o funcionamento do corpo atribuindo um dinamismo vivo e natural a criaturas com mentes como as nossas.

Referências

CARRUTHERS, G. Types of body representation and the sense of embodiment. *Consciousness and cognition*, Elsevier, v. 17, n. 4, p. 1302–1316, 2008.

HANNA, R.; MAIESE, M. *Embodied minds in action*. [S.l.]: OUP Oxford, 2009.

[5] Maiese argumenta que nossa vida e nossos e corpos vivos, não apenas nossos cérebros, ressoam com as outras pessoas; e explora como os problemas da emoção poderiam estar implicados centralmente na esquizofrenia, no autismo e nas psicopatias. Conforme (MAIESE, 2011), na anosognosia uma doença altamente incomum caracterizada por uma incapacidade da pessoa de reconhecer que uma parte do corpo está paralisada. Por exemplo: uma pessoa que realmente não pudesse cuidar de sua vida, de suas experiências ou atividades num determinado tempo não somente teria um lapso no sentido de coerência do seu *self* por um determinado tempo, mas também e totalmente de uma vida consciente como a nossa.

HANNA, R.; THOMPSON, E. Neurophenomenology and the spontaneity of consciousness. *Canadian journal of Philosophy*, Taylor & Francis, v. 33, n. sup1, p. 133–162, 2003.

MAIESE, M. *Embodiment, emotion, and cognition*. [S.l.]: Palgrave Macmillan, 2011.

THOMPSON, E. *Mind in life: Biology, phenomenology, and the sciences of mind*. [S.l.]: Harvard University Press, 2007.

VARELA, F. J.; THOMPSON, E.; RORSCH, E. *The embodied mind: cognitive science and human experience*. [S.l.]: Cambridge, MA: MIT Press, 1991.

A IMPORTÂNCIA DA ESFERA TRANSCENDENTAL NO PENSAMENTO LÓGICO E METAFÍSICO DE WITTGENSTEIN

Ana Claudia Archanjo Veloso Rocha *

Resumo: O presente artigo tem por objetivo apresentar as relações pensadas pelo filósofo austríaco Ludwig Wittgenstein, no que tange a esfera transcendental: lógica e metafísica. Em especial entre os anos de 1914-1929. Para isso, trabalharemos com alguns conceitos fundamentais dentro do pensamento filosófico do autor como "dizer", "mostrar", "limites da linguagem", entre outros. Para tanto nossos fundamentos bibliográficos serão os Notebooks 1914-1916; Tractatus Logico-Philosophicus; Conferência sobre Ética.

Palavras-chave: Linguagem. Lógica. Metafísica. Wittgenstein.

Abstract: This article aims to present the relations conceived by the Austrian philosopher Ludwig Wittgenstein, regarding the transcendental sphere: logic and metaphysics. Especially between the years 1914-1929. For this, we will work with some fundamental concepts within the philosophical thought of the author as "say", "show", "limits of language" , among others. For both our bibliographic will Notebooks 1914-1916; Tractatus Logico-Philosophicus; Lecture on Ethics

Keywords: Language. Logic. Metaphysics. Wittgenstein.

*Professora do Departamento de Filosofia da Universidade Estadual de Montes Claros (Unimontes/MG). Mestre em Filosofia pela Faculdade Jesuíta de Filosofia e Teologia (Faje).

Introdução

O filósofo austríaco Ludwig Wittgenstein é considerado um dos maiores pensadores do século XX. A complexidade de seu pensamento está associada, sobremaneira, às discussões abordadas pelo filósofo que vão desde os assuntos da filosofia analítica aos temas relacionados ao transcendental. A princípio, esta relação parece-nos incompatível, porém a leitura atenta e minuciosa de seus escritos nos aproxima da totalidade do seu constructo filosófico.

Para o pensador, o âmbito do mostrar, ou seja, do transcendental cinge tanto a lógica quanto o metafísico. Isto ocorre porque ambas as esferas possuem elementos transcendentais e que, portanto estão no limite do mundo.

Apesar de serem dois tipos diferentes de aspecto transcendental, observamos que tanto o lógico quanto o metafísico partilham dessa esfera. Ressaltamos que

> A questão de como a lógica pode representar o mundo, e a questão do sentido do mundo, constituem em conjunto o "místico". Ambas são esferas em que as proposições não têm qualquer possibilidade de ser significativas. Assim, a noção de "mostrar" tem raízes em duas relações: a relação entre o mundo e a lógica, e a relação entre os fatos que constituem o mundo e o sentido, ou significado do mundo. (JANIK; TOULMIN, 1991, p. 223).

A afirmação de Bertrand Russell: "Todo problema filosófico, quando submetido à análise e justificação necessárias, revela-se não ser realmente um problema filosófico, ou ser, no sentido em que usamos a palavra, um problema lógico" (Our Knowledge of the external world)(MARCONDES, 2004, p. 53). Entendemos ser pertinente este pensamento de Russell porque segundo Wittgenstein: um problema filosófico é um problema lógico e portanto analítico.

Sendo assim, nosso filósofo em questão, procedeu da mesma forma: o desenvolvimento do pensamento de Wittgenstein parte de uma questão tipicamente analítica, a saber, "qual é o limite da linguagem?". Wittgenstein busca a resposta para tal questionamento, a partir da observação acerca das possibilidades da linguagem. Tais possibilidades, segundo Danilo Marcondes, correspondem ao "teste empírico, adotado como critério de validade das teorias científicas"(MARCONDES, 2004, p. 12).

Observamos que no dia 22 de janeiro de 1915 – período em que o *Tractatus* estava em fase de elaboração – Wittgenstein deixou registrado nos Cadernos 1914-1916 o objetivo de sua escrita: "Toda a minha tarefa consiste em esclarecer a essência da proposição. Quer dizer, indicar a natureza de todos os fatos cuja imagem é a proposição. Indicar a natureza de todo o ser" (WITTGENSTEIN; PROENÇA, 2004, p. 60). Esta declaração nos revela o claro indício de uma busca analítica, quando o filósofo fala acerca de fatos e proposição, todavia, revela um ângulo de possíveis respostas que se aproximam do transcendental quando fala da essência e da natureza de todo o ser.

O que está em observação, conforme já mencionamos são os limites da linguagem. Sendo assim, observamos que a filosofia pensada e desenvolvida acerca de tais limites relaciona-se com a *linguagem científica*. Para isto, a linguagem utilizada precisa ser fundamentada de maneira lógica e em bases empíricas, a fim de que possa ser verificada. Nesse sentido, elementos metafísicos fogem à circunscrição científica, conforme veremos adiante.

A complexa resposta do filósofo relacionada aos limites da linguagem estabelece a fronteira entre aquilo que é possível dizer e, por conseguinte pensar, e aquilo que se mostra.

Wittgenstein afirma no *Tractatus* que "o pensamento é a proposição com sentido" (WITTGENSTEIN, 2008, Aforismo 4). Com efeito, o pensamento a que Wittgenstein se refere é aquilo que se expressa pela linguagem. Conforme mencionado, o pensamento é a proposição dotada de sentido, ele é conectado através da forma lógica ao fato do mundo que está descrevendo. A estrutura lógica, conforme veremos, se mostra na proposição.

Ao alcançar a compreensão de que o que se mostra não é dizível, Wittgenstein vislumbra a esfera do metafísico. É nesse ponto que o degrau analítico é superado e faz-se necessário a chave da perspectiva transcendental para a assimilação e compreensão entre aquilo que é possível ser dito e aquilo que se mostra. Sendo assim, as vertentes *metafísica e lógica* não são excludentes, mas complementares para a total apreensão do pensamento de Wittgenstein.

Isto ocorre devido ao caráter transcendental alcançado pelo filósofo em ambas as esferas do seu pensamento: metafísica e lógica. Embora Wittgenstein tenha como ponto de partida a linguagem que é um fenômeno empírico, ele fundamenta as possibilidades e limites desta, a partir da esfera transcendental.

Esfera lógica

Nosso filósofo coloca o *dizer* e o *mostrar* numa relação peculiar de complementaridade. Ressaltamos que segundo Wittgenstein, o "mostrar" relaciona-se com a esfera transcendental. A linguagem *diz* o que pode ser dito e, ao mesmo tempo, *mostra* as condições transcendentais de possibilidade do dizer. E aquilo que se mostra não pode ser dito.

Quanto à lógica, vejamos uma parte do aforismo 4.121 "A proposição mostra a forma lógica da linguagem. Ela a exibe" (WITTGENSTEIN, 2008, Aforismo 4.121). Notamos que há duas esferas no respectivo aforismo: o dizível, já que o filósofo fala da proposição, mas paradoxalmente é por intermédio da própria proposição que a forma lógica da linguagem não é dizível, mas sim mostrada.

Eis a questão: qual é o domínio do dizer? Só há possibilidade de falar sobre fatos e a linguagem é o instrumento adequado para expressar o que compõe o mundo. Sendo assim, só pode haver proposições dotadas de sentido quando se trata dos fatos mundanos.

No aforismo tractatiano 2.16, Wittgenstein nos diz que: "Os fatos, para serem figuração, devem ter algo em comum com o que é afigurado". Assim os fatos somente representam os elementos do mundo e por conseguinte o ordinário. Nesse caso, observemos que na afirmação tractatiana 2.18: "cada figuração, de forma qualquer, deve sempre ter em comum com a realidade para poder afigurá-la em geral – correta ou falsamente – é a forma lógica, isto é, a forma da realidade". O filósofo associa a figuração à forma lógica, assim, podemos antever que os fatos expressam também a forma lógica da proposição.

Apenas a proposição pode possuir sentido e, quando o possui, é porque ela está afigurando um fato. A única coisa que a proposição contém *a priori* é o seu sentido, pois ela pode ser compreendida antes que saibamos *a posteriori* se é verdadeira ou falsa.

A forma lógica pode ser considerada, na expressão de Wittgenstein, como o "cimento comum à linguagem e ao mundo" (WITTGENSTEIN, 2008, Aforismo 5.4711). Isto se justifica porque a forma lógica é aquela capaz de estruturar o mundo e a linguagem. Esta estruturação só acontece porque é através da lógica que as condições de possibilidade transcendental tanto da linguagem quanto do mundo podem ser estabelecidas. Depois de delimitadas as referidas condições de possibilidade, Wittgenstein chega à essência do mundo. Assim, compreendemos que a lógica é o fundamento tanto da linguagem como do mundo. A princípio, o mais importante é fundamentar os limites da linguagem, mas, quando estes são estabelecidos, necessariamente estabelecem-se as condições de possibilidade do mundo.

Nesse sentido percebemos que a compreensão do Tractatus é lógico-ontológica. Assim, a forma lógica da proposição expressa a condição *a priori* de que as proposições complexas são combinações lógicas de proposições atômicas, descrevendo assim os fatos complexos do mundo, sendo estes combinações lógicas de fatos atômicos. O espaço lógico funciona como a região no interior da qual as proposições atômicas e os fatos atômicos se combinam de acordo com as determinações da forma lógica.

A essência do mundo, ou melhor, de toda a realidade, é fundamentada e estruturada pela forma lógica. Quando Wittgenstein se refere à essência do mundo, ele se refere às condições de possibilidade dos fatos que estão no mundo. Assim, lemos no *Tractatus* que: "Especificar a essência da proposição significa especificar a essência de toda descrição e, portanto, a essência do mundo" (WITTGENSTEIN; PROENÇA, 2004, p. 192). Isto significa que a essência da proposição equivale à essência do mundo, já que os fatos – componentes de toda a realidade – são descritos por proposições. Sob esse aspecto, apresenta-se o plano da ontologia.

É a lógica que fornece a estrutura do nosso pensamento:

> Não podemos pensar nada de ilógico, porque, do contrario, deveríamos pensar ilogicamente. (WITTGENSTEIN, 2008, Aforismo 3.03)
>
> Já foi dito que Deus poderia criar tudo, salvo o que contrariasse as leis lógicas. – É que não seriamos capazes de *dizer* como pareceria um mundo "ilógico". (WITTGENSTEIN, 2008, Aforismo 3.031)

Não há a menor possibilidade de pensar fora do esquadro da estrutura lógica. Nos *Cadernos 1914-1915*, Wittgenstein expressa, no dia 12.09.1916: "torna-se agora claro porque pensei eu que pensar e falar eram o mesmo. O pensar é uma espécie da linguagem. O pensamento é, decerto, *também*[1] uma imagem lógica da proposição e, assim, é igualmente uma espécie da proposição"(WITTGENSTEIN; PROENÇA, 2004, p. 122). Todavia, após refletir melhor sobre a questão, Wittgenstein afirma nos seus escritos dos Notebooks 1914-1916 que "Toda a proposição que tem um sentido tem um sentido completo, e é uma imagem da realidade, de modo que aquilo que nela ainda não está dito não pode simplesmente pertencer ao seu sentido" (WITTGENSTEIN; PROENÇA, 2004, p. 92). Nesse aspecto, podemos entender que se o pensamento é uma imagem lógica da proposição, então poderia o pensamento possuir um sentido completo, já que o pensamento é uma imagem lógica e a proposição é uma imagem da realidade, assim o pensamento pode ser compreendido como uma imagem da realidade.

Apesar da lógica se mostrar nas estruturas do mundo, fazendo-se refletir no empírico, a natureza dela é transcendental. Eis o aforismo tractatiano: "6.13 A lógica não é uma teoria, mas uma imagem especular do mundo. A lógica é transcendental"(WITTGENSTEIN, 2008, Aforismo 6.13). Ela fornece as condições transcendentais de possibilidade da linguagem descritiva, no entanto essa mesma linguagem não pode falar sobre a forma lógica. A lógica estrutura o espaço lógico dentro do qual se inserem os fatos mundanos: ela é transcendental, porque envolve as condições de possibilidade dos fatos que constituem o mundo.

[1] Grifo do autor.

Segundo Margutti:

> Daí o apelo ao mostrar lógico, que revela algo presente na proposição dotada de sentido, como sua condição de possibilidade, mas que não pode ser descrito através de uma proposição. Isto significa postular que aquilo que se mostra logicamente só pode ser contemplado pelo sujeito transcendental, mas numa perspectiva diferente do mostrar místico (MARGUTTI, 2006, p. 30).

O transcendental quando remete-se ao âmbito do lógico é inerente à linguagem e ao mundo, constituindo a essência de ambos e mostrando-se nas proposições e nos fatos que elas descrevem. Vejamos o aforismo tractatiano "4.022 A proposição mostra seu sentido". Nesta afirmação, percebemos que para Wittgenstein o sentido da proposição não é um fato do mundo. Este sentido revela-se na proposição através dos fatos expressos pela proposição.

Dentre as funções da proposição, encontramos uma de suma importância, uma vez que ela é capaz de mostrar a essência do mundo, sem, contudo, ser capaz de dizê-la. A essência do mundo mostra-se pela forma lógica das proposições.

A lógica é necessária e não pode ser em hipótese alguma, uma convenção. A lógica é uma rede transcendentalmente estruturante que o sujeito metafísico utiliza para organizar o mundo dos fatos. Essa rede é necessária no sentido de que é essencial, de que nenhuma outra rede possível seria capaz de realizar essa tarefa. Sendo assim, Wittgenstein concebe a lógica como um grande espelho. Eis as palavras do filósofo no aforismo "5.511: Como pode a lógica, que abrange tudo e espelha o mundo valer-se de sinuosidades e manipulações tão especiais? Só porque tudo isso se entrelaça numa rede infinitamente fina, no grande espelho". Desta forma, entendemos que a lógica reflete as condições transcendentais de possibilidade do mundo. A estrutura lógica se mostra na proposição.

Esfera Metafísica

E quanto a esfera do metafísico? Não há possibilidade de utilização da linguagem para expressar aquilo que se mostra, haja vista que o âmbito do transcendental é o que está no limite do mundo. Destarte, podemos inferir que o mostrável não é um fato, por esse motivo a linguagem não alcança esta esfera. O que podemos entender mediante este pensamento do filósofo? Que aquilo que se mostra independe da vontade do indivíduo, não é do seu domínio. Isto é uma característica complexa da respectiva esfera.

Vejamos o aforismo tractatiano 4.1212 "O que pode ser mostrado não pode ser dito"(WITTGENSTEIN, 2008, Aforismo 4.1212). Podemos perceber a existência de duas dimensões distintas no pensamento do filósofo.

Ao que tudo indica o *mostrar* está no nível transcendental: não se trata aqui de um mostrar empírico, mas de algo que o ultrapassa. Aquilo que se mostra faz parte do místico, do inefável, impossível de ser expresso através da linguagem.

O metafísico, para o filósofo, se remete ao âmbito do indizível. Isto ocorre devido à sua natureza.

Na *Conferência sobre Ética*, Wittgenstein reconhece que, ao procurar pela forma lógica correta das expressões éticas e religiosas, a saber: expressões metafísicas, de repente percebe, num átimo, em silêncio, que não existe forma lógica adequada para as mesmas e que isso constitui uma deficiência essencial da linguagem. Não se pode pretender que um instrumento natural como ela expresse uma essência sobrenatural. Com efeito, uma das propostas de Wittgenstein, ao proferir a *Conferência sobre Ética*, é ressaltar os maus usos das expressões consideradas metafísicas, que pretendem "*ir além do mundo*, o que é o mesmo que ir além da linguagem significativa" (WITTGENSTEIN, 1991, p. 224). Isto implica em ir além do discurso científico, tendo em vista que o metafísico não se encaixa neste tipo de discurso, porque seus objetos de estudo não são fatos.

Tendo em vista a abordagem da esfera transcendental faz-se necessário nos remetermos à impossibilidade de se falar a respeito da metafísica:

> 6.53 O método correto da filosofia seria propriamente este: nada dizer, senão o que se pode dizer; portanto, proposições da ciência natural – portanto, algo que nada tem a ver com filosofia; e então, sempre que alguém pretendesse dizer algo de metafísico, mostrar-lhe que não conferiu significado a certos sinais em suas proposições. Esse método seria, para ele, insatisfatório – não teria a sensação de que lhe estivéssemos ensinando filosofia; mas *esse* seria o único rigorosamente correto. (WITTGENSTEIN, 2008, p. 281)

A metafísica é fundamentalmente transcendental. Logo, seus elementos não podem ser verdadeiros, nem falsos. São simplesmente sem sentido, por não possuírem a possibilidade de verificação. Eis a colocação de Pears sobre este assunto: : "as teorias metafísicas nunca são verdadeiras. São tentativas de dizer aquilo que pode ser mostrado, mas não pode ser dito e somente o que pode ser dito pode ser verdadeiro" (PEARS, 1973, p. 39). Lembramos que o metafísico para o filósofo, embora sem sentido não é sem valor, haja vista que o valor está na mesma dimensão da metafísica, a saber, transcendental. Assim sendo, remetemo-nos ao aforismo tractatiano 6.4, onde o filósofo diz que "Todas as proposições tem igual valor", ou seja, nenhum valor.

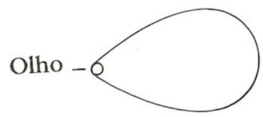
Nesse caminho filosófico, faz-se necessário algumas palavras acerca do sujeito metafísico, também denominado "eu filosófico". A relação entre o sujeito metafísico e aquilo que se mostra envolve uma espécie de intuição ou contemplação.

É o sujeito metafísico que constitui as condições de possibilidade de falar através de um espaço lógico *a priori* que ele impõe ao mundo. Segundo Margutti, tal espaço parece corresponder a um sistema de coordenadas transcendentais, que, infelizmente, não teremos condições de detalhar aqui. De qualquer modo, fica claro que, embora o sujeito metafísico tenha constituído as condições de possibilidade da linguagem e do mundo, ele mesmo não faz parte do mundo.

Eis as palavras do filósofo no aforismo 5.631

> O sujeito que pensa, representa, não existe.
>
> Se eu escrevesse um livro O Mundo tal como o Entendo, nele teria que incluir também um relato sobre meu corpo, e dizer quais membros se submetem à minha vontade e quais não, etc. ? este é bem um método para isolar o sujeito, ou melhor, para mostrar que, num sentido importante, não há sujeito algum: só dele não se poderia falar neste livro.
> O sujeito não pertence ao mundo, mas é um limite do mundo.
> Onde no mundo se há de notar um sujeito metafísico?
> Você diz que tudo se passa aqui como no caso do olho e do campo visual. Mas o olho você realmente não vê. E nada no campo visual permite concluir que é visto a partir de um olho. (WITTGENSTEIN, 2008, p.241-247)

Wittgenstein está se referindo ao sujeito metafísico. Ele não existe "no" mundo, porque esse último não pertence à sua esfera. Por estar no limite do mundo, ele pode contemplá-lo em sua totalidade, *sub specie aeternitatis*. Ele "vê" o mundo sem estar no mundo. Assim como acontece na metáfora do olho e do campo visual:



Fonte: (WITTGENSTEIN, 2008, p. 247)

O que se "vê" sem estar no mundo, é o que se revela por estar em seu limite. Para esta percepção é necessário um elo, um elemento basilar que estabelece uma ponte entre aquilo que se mostra eticamente e a apreensão desta pelo indivíduo. Este elemento é a experiência mística. Mas o que isto vem a ser? É uma experiência pessoal e intrasferível. Como se fosse um súbito instante no qual a esfera mística se mostra. Não há receita ou regras para se alcançar a experiência mística. Ela simplesmente acontece.

É a experiência mística que permite vislumbrar aquilo que se mostra eticamente. A questão da natureza da linguagem e por consequência do que se pode ou não fazer com ela, a questão, em termos tractatianos, do que é dizível e do que é mostrável, nos remete à questão dos limites do mundo. Estes podem ser alterados para aquele que foi agraciado por uma experiência mística. Não é uma alteração física, mas uma modificação da perspectiva de mundo, de entendimento do todo.

Os elementos que compõem tal experiência não são passiveis de descrição. Isso significa que, se tentarmos descrever aquilo que se denomina *experiência mística*, constataremos que não há na linguagem recursos capazes de realizar essa tarefa, uma vez que os elementos que compõem tal experiência não possuem caráter fatual e não podem, portanto, ser descritos.

Tratando-se do aspecto transcendental não seria possível deixar de citar a impossibilidade de proposições religiosas. A inferência de Wittgenstein acerca das ilusórias proposições religiosas foi apontada na *Conferência sobre Ética*. Neste escrito, o objetivo do filósofo foi demonstrar o contrassenso que sobrevém quando se recorre à linguagem para expressar o inefável. Para tanto, ressalta o mau uso do símile e de expressões que parecem adequadas para exprimir o religioso.

A interpretação wittgensteiniana acerca do religioso é que estes são metafísicos e transcendentais, logo, se mostram. A transcendentalidade dos termos religiosos é o grande obstáculo para a elaboração de uma proposição religiosa.

Levando em consideração a impossibilidade de um discurso metafísico, de proposições éticas ou religiosas, enfrentamos uma indagação: qual é a forma adequada para expressar aquilo que se mostra no âmbito do ético? Somente através do silêncio. O silêncio é a única saída, revelando-se como uma espécie de saída contemplativa.

O silêncio proposto por Wittgenstein é referente àquilo que é inexequível de ser dito cientificamente por estar relacionado à mostração do transcendental. O que pode ser dito pertence ao âmbito das proposições da ciência natural, que descrevem fatos. Assim, sempre que alguém tentar dizer algo pertencente à esfera transcendental, ou seja, algo de caráter metafísico, será preciso mostrar-lhe que está tentando falar sobre algo que não pertence ao mundo e que essa tarefa é impossível. A tentativa de falar sobre a esfera transcendental sempre envolve um desrespeito à lógica da linguagem, que pode ser caracterizado pelo uso de um signo sem o correspondente significado. Isso faz com que o método correto em filosofia seja justamente mostrar que certa proposição metafísica não passa de contrassenso, porque deixou de atribuir significado a algum de seus elementos constitutivos.

Com efeito, se o que se mostra é indizível, resta ao sujeito que foi agraciado pela mostração, reter-se em silêncio, permanecendo em profunda imersão no que foi contemplado e em paz consigo mesmo. Este se encontra sem poder nem querer dizer nada sobre o assunto, por saber que não seria capaz de relatar algo oriundo de uma experiência tão profunda e pessoa[2]. Eis a 7ª proposição: "O que não se pode falar, deve-se calar".

É fato que antes de chegar ao estado do silêncio apaziguador, o sujeito empírico enfrenta um conflito consigo mesmo, uma vez que ele tem os instrumentos para expressar o que está no mundo, mas em hipótese alguma consegue expressar-se sobre o que está no limite do mundo, por não ter os instrumentos linguísticos adequados: "Toda a experiência é mundo e não necessita do sujeito" (WITTGENSTEIN, 1998, p.131). O sujeito a que Wittgenstein se refere neste caso é o sujeito metafísico que não está localizado no mundo. Quando a experiência de contemplação acontece, é possível constatar que ela pertence ao sujeito metafísico, que não se encontra no mundo, mas no limite deste.

Conclusão

Concluímos assim que a importância da esfera transcendental no pensamento de Wittgenstein é extremamente relevante para a constituição da sua filosofia. Tal esfera abrange os dois aspectos mais importantes de sua critica à linguagem, a saber: o lógico e a metafísico. Tanto a esfera lógica quanto a ética estão relacionadas à esfera mística. Embora sejam de aspectos diferenciados. No entanto, é interessante observar que aquilo que se "mostra" , não é factual. Assim sendo, pensar em proposições éticas, religiosas, estéticas ou que expressem a essência lógica da linguagem não são possíveis.

Sendo assim, qualquer tentativa de expressar o transcendental é sem sentido por não haver fatos. Destarte, não pode haver proposições dotadas de sentido quando se trata do que está no limite do mundo, ou seja, na esfera transcendental.

A falta de sentido não é porque ainda não se encontrou uma forma adequada de se expressar, mas é em função do sentido estar ausente na natureza dos temas em questão: "Em outras

[2] Não podemos nem chamar de experiência por ser da ordem do sujeito metafísico, mas por falta de vocabulário, utilizarei o termo em questão.

palavras, vejo agora que essas expressões carentes de sentido não careciam de sentido por não se ter ainda encontrado as expressões corretas, mas sua falta de sentido constituía sua própria essência"(WITTGENSTEIN, 1995, p. 224). Como podemos ver, Wittgenstein reconhece que, ao procurar pela forma lógica correta das expressões éticas e religiosas, ele de repente percebe, num átimo, em silêncio, que não existe forma lógica adequada para as mesmas e que isso constitui uma deficiência essencial da linguagem. Não se pode pretender que um instrumento natural como ela expresse uma essência sobrenatural.

Referências

JANIK, A.; TOULMIN, S. *A Viena de Wittgenstein*. Rio de Janeiro: Campus, 1991.

MARCONDES, D. *Filosofia analítica*. Rio de Janeiro: Zahar, 2004.

MARGUTTI, P. R. P. *Iniciação ao silêncio - análise do Tractatus de Wittgenstein*. São Paulo: Edições Loyola, 2006.

PEARS, D. *As ideias de Wittgenstein*. São Paulo: Cultrix, 1973.

PINTO, P. R. O tractatus de wittgenstein como obra de iniciação. *Filosofia Unisinos*, v. 5, n. 8, p. 81–104, 2004.

WITTGENSTEIN, L. *Tractatus logico-philosophicus*. São Paulo: Edusp, 2008.

WITTGENSTEIN, L. *Fichas Zettel*. Portugal: Edições 70, 1989.

WITTGENSTEIN, L. *Aulas e Conversas sobre estética Psicologia e fé religiosa.* Tradução de Miguel Tamen. Lisboa: Edições Cotovia: [s.n.], 1991.

WITTGENSTEIN, L. *Conferência Sobre Ética.* Tradução de Darlei DallAgnol. Florianópolis/SC: Editora da UFSC/Editora UNISINOS: [s.n.], 1995.

WITTGENSTEIN, L. *Cultura e valor*. Portugal: Edições 70, 1998.

WITTGENSTEIN, L. *Luz y sombra: una vivencia (sueño) nocturna y un fragmento epistolar*. [S.l.: s.n.], 2006.

WITTGENSTEIN, L. *Diarios secretos*. [S.l.]: Alianza Editorial, 2008.

WITTGENSTEIN, L. *Movimentos de pensamento: diários de 1930-32/1936-37*. São Paulo: Martins Fontes, 2010.

WITTGENSTEIN, L.; PROENÇA, J. T. *Cadernos: 1914-1916*. [S.l.]: Edições 70, 2004.

WITTGENSTEIN, L. et al. *Cartas a Russell, Keynes y Moore*. [S.l.]: Taurus, 1979.

SABERES - Special Issue

João Daniel Dantas,
Patrick Terrematte,
Sanderson Molick

$\Phi\iota\lambda o\mu\varepsilon\nu\alpha$ II
(FILOMENA 2)

The FILOMENA Workshop (Philosophy, Logic and Metaphysics), promoted by the **Group on Logic and Formal Philosophy** from the UFRN, has the purpose of gathering logicians working at the intersection of Logic and Metaphysics, through the application of formal methods in Philosophy.