ARGUMENTOS
Revista de Filosofia

ISSN 1984-4247

Special Issue:

BRENTANO
and Philosophy of Mind

Ano 7
N°. 13
2015

Revista do Programa de Pós-Graduação em Filosofia da Universidade Federal do Ceará

ARGUMENTOS

Revista de Filosofia

Revista do Programa de Pós-Graduação em Filosofia da
Universidade Federal do Ceará - UFC

Argumentos

Revista de Filosofia

EDIÇÃO
Coordenação Editorial: Odílio Alves Aguiar
Projeto Gráfico, Editoração e Capa: Sandro Vasconcellos
Imagem da Capa: Franz Brentano - https://upload.wikimedia.org/wikipedia/commons/3/30/Franz_Brentano1.png
Bibliotecária: Perpétua Socorro T. Guimarães - CRB 3/801

ENDEREÇO PARA CORRESPONDÊNCIA
Campus do Pici - Instituto de Cultura e Arte (ICA)
Fortaleza - CE - CEP 60455-760
Site: www.filosofia.ufc.br/argumentos

E-mail: argumentos@ufc.br
Solicita-se permuta

Periodicidade: Semestral
Ano 7 - Número 13 - Fortaleza, jan./jun. - 2015
ISSN: 1984-4247

Dados Internacionais de Catalogação na Fonte
Bibliotecária: Perpétua Socorro Tavares Guimarães - CRB 3/801

Argumentos - Revista de Filosofia - 2015
Fortaleza, Universidade Federal do Ceará – Programa de Pós-graduação em Filosofia, ano 7, n. 13, semestral, jan./jun. 2015.

1. Filosofia I. Universidade Federal do Ceará

CDD: 100

ISSN: 1984-4247

Sumário

VARIA:

RESENHA:

Presentation

Intentionality and Consciousness was the main theme of the "Sixth Internacional Coloquium in Philosophy of Mind", organized in Fortaleza, in september 2011. In that occasion, researchers from England, Portugal, Canada, Argentina, Germany, and from different parts of Brazil, gathered to discuss that classical theme from different points of view.

Later on, came out the idea of publishing some results of the conference in a special issue of *Argumentos,* adopting the form of a *disputatio,* with a target paper written by Denis Fisette, comments by a few specialists, and replies by the author. The comments are disposed by alphabetic order of the first name, as we use to do in Brazil. Fisette's paper is about Brentano's legacy in the theory of consciousness and compares Brentano's ideas with those of contemporary authors like David Rosenthal. Interestingly, Fisette's conclusions show that Brentano's theory of consciousness avoid some of the main criticisms suffered by Rosenthal's Higher Order Theory (HOT).

The editors are proud to released this issue of *Argumentos* which constitutes a collective work on a philosopher considered today as the source of two of the most important philosophical movements of the XXth Century: the phenomenological movement and the analytic movement.

It is important to inform that *Argumentos* publishes two issues each year, one on Pratical Philosophy and another one, on Theoretical Philosophy. The present issue is a theoretical one. In its section Varia, we present some other contributions towards this general subject written by philosophers from Brazil and from other nations.

Invited editors
André Leclerc, Marcos Silva

Apresentação

Intencionalidade e consciência foram os temas principais do "VI Colóquio Internacional em Filosofia da Mente", organizado em Fortaleza, em setembro de 2011. Nesta ocasião, pesquisadores da Inglaterra, Portugal, Canadá, Argentina, Alemanha, e de outras partes do Brasil, se reuniram para discutir este clássico tema sob diferentes perpectivas.

Posteriormente, surgiu a idéia de se publicar alguns resultados deste evento em um número especial da *Argumentos*, adotando a forma de uma *disputatio*, com o artigo-alvo escrito por Denis Fisette, comentários de alguns especialistas e réplicas do próprio autor. Os artigos-comentários são dispostos em ordem alfabética pelo primeiro nome do autor, como nós fazemos usualmente no Brasil. O artigo de Fisette trata do legado de Brentano na teoria da consciência e compara as idéias de Brentano com ideias de autores contemporâneos como David Rosenthal. Interessantemente, as conclusões de Fisette mostram que a teoria da consciência de Brentano evita algumas das principais críticas sofridas pela Teoria de Ordem Superior (HOT) de Rosenthal.

Os editores estão orgulhosos por publicarem este número da *Argumentos*. Este constitui um trabalho coletivo sobre um filósofo considerado hoje como a fonte dos dois movimentos filosóficos mais importantes do século XX, a saber: o movimento fenomenológico e o movimento analítico.

É importante informar que a *Argumentos* publica todo ano dois números, um em Filosofia Prática, e outro, em Filosofia Teórica. O presente número corresponde ao último. Na seção Varia, apresentamos outras contribuições sobre este tópico geral escritas por filósofos do Brasil e de outras nações.

Editores convidados
André Leclerc, Marcos Silva

ARGUMENTOS
Revista de Filosofia

Denis Fisette*

Target paper: Franz Brentano and higher-order theories of consciousness[1]

ABSTRACT

This article addresses the recent reception of Franz Brentano's writings on consciousness. I am particularly interested in the connection established between Brentano's theory of consciousness and higher-order theories of consciousness and, more specifically, the theory proposed by David Rosenthal. My working hypothesis is that despite the many similarities that can be established with Rosenthal's philosophy of mind, Brentano's theory of consciousness differs in many respects from higher-order theories of consciousness and avoids most of the criticisms generally directed to them. This article is divided into eight parts. The first two sections expound the basic outline of Rosenthal's theory, and the third summarizes the principal objections that Rosenthal addresses to Brentano, which I, then, examine in sections 4 and 5. In sections 6 and 7, I discuss Brentano's principle of the unity of consciousness, and in section 8, I consider the scope of the changes that Brentano brings to his theory of consciousness in his later writings, which follow the 1874 publication of *Psychology*. I then draw the conclusion that Brentano's theory rests on a view of intransitive and intrinsic self-consciousness.

Keywords: Brentano; Higher-order theories; Consciousness; Self-consciousness.

RESUMO

Este artigo trata da recente recepção dos escritos de Franz Brentano sobre a consciência. Estou particularmente interessado na conexão estabelecida entre a teoria da consciência de Brentano e as teorias de ordem superior da consciência e, mais especificamente, na teoria proposta por David Rosenthal. Minha hipótese

* Université du Québec à Montréal. denis.fisette@uqam.ca

[1] A version of this article was presented in September 2011 at the Federal University of Ceará in Fortaleza, Brazil during the Sixth International Congress of Philosophy of Mind. I wish to thank André Leclerc for his comments on a previous version of this article, Denis Courville for his work on the English version of this paper, and the Social Sciences and Humanities Research Council of Canada for its financial support.

de trabalho é que, apesar das muitas similaridades que possam ser estabelecidas com a filosofia da mente de Rosenthal, a teoria da consciência de Brentano difere em muitos aspectos das teorias de ordem superior e evita boa parte das críticas geralmente dirigidas a elas. Este artigo é dividido em oito partes. As primeiras duas seções expõem o arcabouco básico da teoria de Rosenthal, e a terceira resume as principais objeções que Rosenthal dirige a Brentano, que eu, então, examino nas sçcões 4 e 5. Nas seções 6 e 7, discuto o princípio da unidade da consciência de Brentano, na secão 8, considero o alcance das mudanças que Brentano faz em sua teoria da consciência em escritos posteriores à publicação de *Psicologia* em 1984. Eu, então, concluo que a teoria de Brentano repousa sobre a visão de uma auto-consciência intrínseca e intransitiva.

Palavras-chave: Brentano; Teorias de ordem superior; Consciência; Auto-consciência.

The theory of consciousness put forth by Franz Brentano in *Psychology from an Empirical Standpoint*[2] has recently been a topic of interest in the philosophy of mind and cognitive sciences. This growing interest must be understood in connection with the current debates on the so-called "problem of consciousness". This problem, which has been at the center of discussions in philosophy of mind for more than thirty years now, refers to the difficulties of both defining consciousness and explaining it according to the descriptive apparatus that is currently available.[3] This problem is also known, ever since D. Chalmers (1995), as the "hard problem" of consciousness, given the specific challenge of explaining scientifically (phenomenal) consciousness in the context of the cognitive sciences.

Faced with this problem, some philosophers have recently developed theories of consciousness, which follow in some respects in the steps of Brentano's theory of consciousness, thereby emphasizing its relevance and its significance in the context of the recent debates about consciousness[4]. Such is the starting point of a debate on what has come to be known as the neo-Brentanian theories of consciousness. This debate is partly exegetic because it deals with how Brentano's psychology exposed in *Psychology* is to be interpreted. But the main philosophical issue at stake in this debate concerns the viability of Brentano's theory of consciousness with regards to the problem of consciousness.

[2] I will use the abbreviation *Psychology* to refer to the English translation of *Psychologie vom empirischen Standpunkt*, and *Schriften I* for the German edition provided by Ontos. Other abbreviations used in this text are indicated in the bibliography at the end of this article.

[3] See D. Fisette and P. Poirier (2000).

[4] This is particularly the case of Uriah Kriegel who maintains in a recent article entitled "Brentano's Most Striking Thesis" (forthcoming) that Brentano's theory represents currently one of the main options available in philosophy of mind.

I am particularly interested here in an interpretation of Brentano's theory of consciousness which currently prevails in Brentanian studies and which is based on higher-order theories of consciousness[5]. Neo-Brentanians, like most critics of Brentano, share the view that the latter's theory constitutes a version of a higher-order theory of consciousness[6]. Such is also the interpretation of David Rosenthal, one of the most notable supporters of higher-order theories of consciousness, who has emphasized on many occasions the importance of Brentano's contribution to philosophy of mind, most notably in the context of an interpretation of the main principles of the theory of consciousness put forth by Brentano in Book II of *Psychology*[7]. In spite of disagreeing with some of these principles, Rosenthal (1991, p. 30) nevertheless considers that the heart of the Brentanian theory of consciousness "is virtually indistinguishable from that for which [he] argue[s]".

That being said, opinions differ with regards to the significance of Brentano's theory. Critics of Brentano maintain that his philosophy of mind is obsolete in that it conveys the same assumptions as those of higher-order theories of consciousness, all of which were already denounced by most of Brentano's students, most notably by Husserl and his students. Brentano's work on intentional consciousness would therefore be of no use in addressing contemporary issues in philosophy of mind[8]. Dan Zahavi (2004), for example, holds that Brentano and higher-order theories of consciousness cannot adequately account for (self-) consciousness since both fail to distinguish between consciousness and intentionality:

> Any convincing theory of consciousness has to be able to explain the distinction between *intentionality*, which is characterized by an epistemic *difference* between the subject and the object of experience, and *self-consciousness*, which implies some form of *identity*. But this is precisely what

[5] On the connection established between Brentano's theory and higher-order theories of consciousness, see in particular G. Güzeldere (1997, p. 789); C. Siewert (1998, p. 357-358); D. Zahavi (1998, p. 130-131; 2004, p. 73; 2006, p. 7); V. Caston (2002, p. 754); M. Textor (2006, p. 412); G. Janzen (2008); and R. Gennaro (1996, p. 27–29).

[6] There are also other interpretations of Brentano's theory that question this connection, but they nevertheless assume as a starting point the same presupposition as the neo-Brentanian theories of consciousness. See A. Thomasson (2000) and J. Brandl (forthcoming).

[7] Rosenthal comments Brentano's psychology in many of his articles, most notably in D. Rosenthal (2011; 2009; 2005; 2003; 1997; 1993; 1991).

[8] We can mention, for instance, most notably the criticism that Husserl addresses to Brentano in the *Logical Investigations* (D. FISETTE, 2010), which have been echoed by some of the Husserl's students, such as A. Gurwitsch (D. Zahavi, 2006, p. 4) and R. Ingarden (1969). The latter in particular suggests that one should do away with what he considers to be Brentano's main idea, that is, that one can only be conscious of an act through a representation of the said act: "One must rather admit that consciousness and, particularly, the act of consciousness, for example, of perception is something that is lived-through (*Durchleben)*, a certain form of self-knowledge, where there is no need to introduce reflection, representation, or judgment". (R. INGARDEN, 1969, p. 629, my translation) Thus, the debate is a not new one. It is also at the center of a debate which opposes E. Tugendhat (1979), who defends a position similar to Rosenthal's, and the members of the Heidelberg School. (D. ZAHAVI, 1998, p. 130-131).

> the higher-order theory, which seeks to provide an extrinsic and relational account of consciousness, persistently fails to do. (ZAHAVI, 2004, p. 70).

This objections bears a resemblance to what C. Siewert (1998, p. 197), a further critic of Brentano's theory, calls the "conscious-of trap" or what is also known as "intentionalism".[9]

These criticisms presuppose, however, a certain interpretation of Brentano's philosophy of mind that has prevailed within Brentanian studies ever since the publication of R. Chisholm's writings in which he maintains that intentionality is the fundamental concept in Brentano's theory of the mind. Hence what has been termed as "Brentano's thesis," which states that intentionality is what constitutes for Brentano the fundamental characteristic of the mind.[10] Brentano deserves credit for having reintroduced intentionality as a key philosophical notion which still remains significant in the context of contemporary philosophy. However, it is one thing to acknowledge that Brentano has reactualized the notion of intentionality, it is quite another to take it to be the central thesis at the heart of his psychology. For as the recent reception of Brentano's writings has shown, this intentionalist reading rarely takes into consideration the other principles of Brentano's *Psychology* and, particularly, of his theory of consciousness, which represents the central theme of Book II of *Psychology*, where intentionality is introduced.[11] Furthermore, Brentano's writings on the topic of consciousness that follow the publication of *Psychology* in 1874 provide further arguments against the presupposition that underlies this interpretation.

This article addresses the recent reception of Brentano's writings on consciousness. I am particularly interested in the connection established between Brentano's theory of consciousness and higher-order theories of consciousness and, more specifically, the theory proposed by Rosenthal. The latter's remarks on Brentano's theory of consciousness in *Psychology* will serve as this article's common thread. My working hypothesis is that despite the many similarities that can be established with Rosenthal's philosophy of mind, Brentano's theory of consciousness differs in many respects from higher-theories of consciousness and avoids most of the criticisms generally directed at them.[12] I will argue that Brentano's theory rests on a view of intransitive and intrinsic self-consciousness.

---

[9] Intentionalism is the thesis that intentionality is the (only) mark of the mental, or that a conscious mental state is mainly determined by its intentionality. One of the proponents of this thesis is T. Crane who at times similarly attributes it to Brentano (T. CRANE, 2007).

[10] For a criticism of this thesis attributed to Brentano since Chisholm, see D. Moran (1996).

[11] See the papers of J. Brandl, U. Kriegel and M. Textor collected in the first section of *Themes from Brentano* (in D. FISETTE AND G. FRÉCHETTE (Eds.) (2013, p. 23-86)).

[12] See R. van Gulick (2000) for a comprehensive summary of the main points of criticism raised against higher-order theories of consciousness.

Two Concepts of Consciousness

Rosenthal distinguishes between two main traditions at the source of the contemporary trends within philosophy of mind, namely Cartesianism and Aristotelianism. Each tradition exemplifies a view of consciousness which can be identified by combining two fundamental concepts in philosophy of mind, namely intentionality and consciousness. According to the Aristotelian tradition, to which Rosenthal claims to belong, the essential property of the mental is intentionality, and Rosenthal's own theory of consciousness, better known as a higher-order theory of consciousness (hence the acronym HOT), endorses the reduction of consciousness to an intentional relation between a higher-order thought and its object. Within the Cartesian tradition, on the other hand, the mind is characterized by consciousness, and intentionality is thus understood as a mode of relation between consciousness and its objects.[13] Moreover, Rosenthal maintains that the way we understand consciousness is determined by our adherence to either one of these concepts of the mind. Interpreters of Brentano are divided on the question of whether the view of consciousness endorsed by Brentano in *Psychology* makes him a Cartesian or an Aristotelian in the area of philosophy of mind.[14] Before suggesting an answer to this question, we must consider some of the features that Rosenthal attributes to each of these concepts of the mind.

Let us begin with a distinction between two notions of consciousness, namely state consciousness and what Rosenthal calls "creature consciousness", that is, the consciousness of an organism or what could simply be referred to as subjective consciousness. To attribute the predicate of "being conscious" to a state simply means that a mental state has the property of being conscious. For example, a persistent stomach pain may be conscious or not depending on whether we pay attention to it or not. On the other hand, the notion of creature consciousness simply refers to the property that an agent has of being awake

[13] A passage from Rosenthal's classic article "Two Concepts of Consciousness" summarizes well the opposition: "Thus writers with Cartesian leanings have generally favored some mark based on consciousness, while those in a more naturalist, Aristotelian tradition have tended to rely instead on some such mark as intentionality or sensory character" (1986, p. 335). One of Rosenthal's arguments against Cartesianism is that by defining consciousness as an intrinsic property, it deprives us of the possibility of providing a satisfactory (naturalist) explanation of consciousness. (D. ROSENTHAL, 2003, p. 166; 1997, p. 735)

[14] In many of his articles (ROSENTHAL, 1990, p. 746-7; 1991, p. 30; 2004, p. 30 sq.; 1993, p. 211-212; 2009, p. 4), Rosenthal describes Brentano as a Cartesian, but we will later see that many other aspects of the latter's theory of consciousness brings him rather closer to Aristotelianism. It goes without saying that the notions of Cartesianism and Aristotelianism such as they are used by Rosenthal represent first and foremost two general views of consciousness and, to a lesser extent, two historical currents to which these two notions also refer. The influence that Descartes exerted on Brentano's philosophy should not be neglected (D. FISETTE, 2015), but the main inspiration for his theory of consciousness, just as for his ontology, is without any doubt Aristotle, as Brentano himself indicates on many occasions in *Psychology* (V. CASTON, 2002). We should also note that Herman Schell, a student of Brentano, had published in 1873 a doctoral thesis dedicated to the latter on the topic of the unity of consciousness in Aristotle, a fact that is not trivial given that Brentano was very directive with respect to his students' research. (H. SCHELL, 1873).

or, say, of being in a deep coma. By favoring the latter view, which Cartesianism seems to do, a theory of consciousness seems incapable of accounting for what it is for mental states to be conscious other than by stating that an agent is simply conscious (of all his thoughts).[15]

A second distinction that we also owe to Rosenthal refers to two uses of the attribute "being conscious" which figures in the definition of both concepts of consciousness: an intransitive use, which requires no accusative object (such as, for example, to be conscious or unconscious, to be anxious, to be in a good mood or excited, etc.) and a transitive use which makes use of an accusative object (such as, for example, to be conscious of some noise, to be conscious of the fact that returning to class (after the strike) will be difficult, *etc*.). Transitive consciousness is another term meant for intentional consciousness and refers to the relation that an agent has to something:

> One is transitively conscious of something if one is in a mental state whose content pertains to that thing - a thought about the thing, or a sensation of it. That mental state need not be a conscious state (ROSENTHAL, 1997, p. 737).

This notion pertains first and foremost to the subject insofar as one cannot say of a mental state that it is in itself conscious of anything (ROSENTHAL, 1997, p. 738). Used in an intransitive sense, the term "conscious" refers to a monadic predicate that stands as a non-relational property, such as in the definition of subjective consciousness.

The distinction between "being conscious" in an intransitive and a transitive sense is associated with another distinction established between two types of properties ascribable to mental states, namely intrinsic properties and extrinsic properties. The latter distinction finds its linguistic expression in the previous distinction between the transitive and intransitive uses of the predicate "being conscious". Considered as a monadic predicate, it refers to an intrinsic property, while when used as a relation, it characterizes, instead, an extrinsic property:

> A property is intrinsic if something's having it does not consist, even in part, in that thing's bearing some relation to something else. If being conscious is at least partly relational, a mental state could be conscious only if the relevant relation held between the state and some other thing. (ROSENTHAL, 1997, p. 736).

[15] This view of consciousness attributed to Descartes also serves as the starting point of David Armstrong's analysis of consciousness in his book *The Nature of Mind*: "There is, however, one thesis about consciousness that I believe can be confidently rejected: Descartes's doctrine that consciousness is the essence of mentality. That view assumes that we can explain mentality in terms of consciousness. I think that the truth is in fact the other way round. Indeed, in the most interesting sense of the word 'consciousness,' consciousness is the cream on the cake of mentality, a special and sophisticated development of mentality. It is not the cake itself." (D. ARMSTRONG, 1997, p. 721).

We may now formulate, with the help of these terminological distinctions, the concepts of consciousness that correspond respectively to Cartesianism and Aristotelianism. A theory of higher-order thoughts regards consciousness as an extrinsic, transitive and relational property of mental states, that is, as an intentional relation between a higher-order thought and its object. To use the example of a stomach pain, the higher-order thought that accompanies the initial pain state could be expressed as: "I am presently feeling pain in my stomach". A sensory state that would not be accompanied by such a thought could not be, strictly speaking, a pain given that for most higher-order theories of consciousness this sensory quality does not exist prior to the thought or the perception that we have of it. For this pain state to be conscious, we must be transitively conscious "of" this state, and in order to be transitively conscious of it, we must have a higher-order thought about the targeted initial state, thereby making it conscious. This theory rejects Cartesianism insofar as the latter maintains that consciousness is a non-relational, intransitive and intrinsic property of the mind (ROSENTHAL, 1997, p. 737). According to Rosenthal, all of modern philosophy up to Brentano has come to understand consciousness as an intrinsic and intransitive property of agents and it was therefore assumed, for this reason, that the agent was conscious of all his thoughts or mental states. In support of this diagnostic, Rosenthal quotes the passage of the *Meditations* ("Fourth Reply") in which Descartes maintains that "no thought can exist in us of which we are not conscious at the very moment it exists in us" (1964-1965, p. 246; translation from ROSENTHAL, 1997, p. 747). Hence the criticism that Rosenthal opposes to Cartesianism of confusing state consciousness with subjective consciousness, that is, of merging a mental state's being conscious in virtue of which one is *intransitively* conscious of that state with one's being conscious of that state in virtue of which one is *transitively* conscious of being in that state.

That being said, it seems that Brentano, by insisting more on state consciousness than on subjective consciousness while, nevertheless, regarding consciousness as an intrinsic property of mental states, holds a middle position between Cartesianism and Aristotelianism. This is at least the interpretation that Rosenthal has proposed in a recent article, where he maintains that the originality of Brentano's theory of consciousness, in comparison to that of the Cartesian tradition, lies in the thesis that all psychical (or mental) states are conscious (D. ROSENTHAL, 2009, p. 2)[16]. Hence the breakthrough that Brentano's theory represented historically insofar as it provided an explanation "both of what it is for states to be conscious and of why, as he held, all mental

[16] Rosenthal explains later on in the same article: "it was rare until Brentano's time to describe mental states as conscious at all. Even though Descartes and Locke were plainly writing about the property we describe as a state's being conscious, they did not say that our mental states are all conscious, but rather that we are conscious of all our mental states." (ROSENTHAL, 2009, p. 4).

states are conscious" (ROSENTHAL, 2009, p. 2). Part of this explanation lies in Brentano's theory of primary and secondary objects, which I will later discuss.

Rosenthal and higher-order theories of consciousness

Let us first examine precisely how this form of Aristotelianism expresses itself in Rosenthal's theory. This theory shares with other higher-order theories of consciousness many features (R. VAN GULICK, 2000, 2006). As their name indicates, higher-order theories make a distinction between lower-order and higher-order states. Lower-order states may be either qualitative states such as pain and moods or intentional states such as desire, belief, etc. However, many of these theories maintain that these two types of states are numerically distinct in the sense that they exist independently of one another. Conscious states are also distinguished from non-conscious states; a non-conscious state consists in a higher-order state, which is by definition not accompanied by a higher-order state that would make it conscious. The postulate that there are non-conscious mental states is common to all higher-order theories, and it raises many questions when considered in relation to the issue of *qualia* (is it possible, for example, for one to feel pain without being conscious of it?). Thus, a conscious state is a state accompanied by a higher-order state (or a meta-state). To have a pain, for example, presupposes a higher-order perception or thought of the type: "I presently have or feel a pain"; to have the desire to eat seafood or to have inclinations towards abstract things assumes a meta-state of the type: "I presently have the desire for or the inclination towards something". This meta-state is intentional; it is about a lower-order state which it targets. Given that consciousness is for many of these theories a relational and extrinsic predicate, it is the intentional relation between the higher-order state and the target state that makes the latter conscious. However, the conscious state must be immediate and non-inferential. In other words, the process by which the higher-order perception or thought bears a relation to the initial state is not itself conscious. Lastly, these theories all insist on the reflexive character of the content of the higher-order mental state.

That being said, there are significant differences between the various versions of higher-order theories, the most important being what distinguishes Armstrong's theory (higher-order perception or HOP) from Rosenthal's. They differ first and foremost on the question of the psychological mode of the higher-order state (whether a thought or a perception) and on the role played by introspection. Rosenthal unequivocally rejects the perceptual model upheld by Armstrong and, more recently, by W. Lycan on the basis that there is, on the one hand, no empirical support for the existence of a monitoring consciousness as held by HOP and, on the other hand, that higher-order thoughts, in contrast to perception, lack any qualitative properties (ROSENTHAL, 2005, p. 105-109).

Moreover, Rosenthal suggests that the concept of introspection must be revised as to insist on the fact that it is independent of qualities and of perceptual monitoring, as I will later further discuss.

What is specific to Rosenthal's higher-order theory is the role that it assigns to thoughts and, more precisely, to contents of propositional attitude and the relation that these higher-order thoughts bear to their target states. Returning to the example of stomach pains, we may express the higher-order thought that accompanies such initial states in the following way: "I now have or (feel) a pain in my stomach". A sensory state that would not be accompanied by a thought of this type would not be, strictly speaking, a pain because, as we have already indicated, this sensory quality does not exist prior to the thought that we have about it. In order for this pain state to be conscious, we must be transitively conscious of it, and to be transitively conscious of such a state means that we have a higher-order thought about it, such that it makes the latter conscious. This is the central thesis of Rosenthal's theory, which he succinctly summarizes as follows:

> We are conscious of something, on this model, when we have a thought about it. So a mental state will be conscious if it is accompanied by a thought about that states. […] The core of the theory, then, is that a mental state is a conscious state when, and only when, it is accompanied by a suitable HOT. (ROSENTHAL, 1997, p. 741).

The heart of this theory may be reformulated with the help of the following definition: a mental state *M* of a subject *S* is conscious iff *S* has another mental state, *M**, in such a way that *M** is an appropriate representation of *M*. As in many of these theories, *M* refers here to the target states which are either intentional, such as in the intention of planning a trip, or non-intentional, such as in a pain or in the aesthetic pleasure taken in a work of art. *M** refers to a belief state whose assertive modality and whose content makes the target state conscious. What is thus meant by "appropriate representation of *M*" is that *M** is an assertive state which, strictly speaking, can be the only state to perform such a function given that it is by means of this belief that the agent posits the existence of the target state, and thereby becomes conscious of it. A doubt, a desire or any other state that does not have this quality or this mode may not adequately perform such a function.

One of the fundamental principles accepted by any higher-order theory of consciousness is the *transitivity principle*, which Rosenthal defines as consisting in "the view that a state's being conscious consists in one's being conscious of that state". (ROSENTHAL, 2009, p. 4; see also 2005, p. 4). As Rosenthal indicates, this principle imposes a new constraint on the specific content of any higher-order thought, namely that "one is, oneself, in that very mental state". (ROSENTHAL, 1997, p. 740-741). To be conscious consists in

being, oneself, in a given mental state, which is not the same thing as being conscious of our mental states, as maintained by Cartesianism. For all conscious states are my own states and first-person accessible: I can only be conscious of my own stomach pain and not someone else's. Rosenthal follows Aristotle and Brentano, who maintain that when the subject perceives, believes or desires something, she is conscious not only of what she perceives, believes or desires, but also of being in these states or of performing these acts. But, contrarily to Brentano and Aristotle, Rosenthal argues that higher-order thoughts are unconscious in that we generally do not notice that we are aware of being in such states. Hence the appeal to a third-order thought to account for the process by which one becomes explicitly aware of the content of the state that one is in:

> A mental state is conscious only if it is accompanied by a HOT. So that HOT will not itself be a conscious thought unless one also has a third-order thought about the second-order thought. (ROSENTHAL, 1997, p. 742).

By postulating third-order thoughts, Rosenthal is, then, able to account for introspection. But introspection should not be understood, as in Armstrong's model of consciousness, as a perception or an internal monitoring mechanism. Rosenthal conceives of introspection rather in reference to attention and by means of the opposition between focal and peripheral consciousness:

> A state is introspectively conscious only when one is conscious of it in an attentive, deliberate, focused way, whereas states are non-introspectively conscious when our awareness of them is relativity casual, fleeting, diffuse, and inattentive. (ROSENTHAL, 2005, p. 107).

The notion of introspection put forth by Rosenthal is therefore very different from that which is criticized by Brentano in his *Psychology*, and such a notion is actually not too remote from Brentano's own notion of inner perception as we will later see.

Brentano's intrinsicalism and the self-representational theory of consciousness

Let us now turn to Rosenthal's reading of Brentano's theory of consciousness. One immediately remarks Rosenthal's insistence on the aspects of Brentano's theory that differ from his own more than the aspects which bring it closer to a higher-order theory of consciousness. First with respect to some of the similarities, we should note that Brentano, like many higher-order theories of consciousness, makes a distinction within his classification of mental states between lower-order states (representations) and higher-order states (judgment

and emotions). Furthermore, Brentano's notion of judgment (or belief) performs a function similar to that assigned to higher-order thoughts by Rosenthal. Indeed, Brentano regards it as a mode of consciousness and as a relational property of mental states[17]. But there are also significant differences between both theories of consciousness; the main two being the unconscious character of higher-order thoughts and the thesis that consciousness is an extrinsic property of mental states. The main point of contention between Rosenthal and Brentano concerns the question whether consciousness is ultimately an intrinsic or an extrinsic property of mental states[18]. There are three main problems associated with the view of intrinsicalism, which Rosenthal attributes to Brentano and which is of particular interest in the context of our analysis. The first concerns the infinite regress objection, which Brentano discusses at length in *Psychology* in connection with the hypothesis of the existence of unconscious mental states. The problem faced by Brentano's theory is that by rejecting this hypothesis he must explain how the thesis that all mental states are intrinsically conscious does not culminate in an infinite regress. The second problem refers to what van Gulick has termed the "distinctness assumption", that is, the thesis that higher-order and lower-order states are numerically distinct. This problem addresses the relation that Brentano establishes between target states (for example, the representation of a sound) and higher-order states (for example, the judgment about the represented sound). The third problem faced by intrinsicalism is that of the individuation of mental states.[19]

Before we discuss these objections, we should take note of a certain ambivalence on Rosenthal's part in his interpretation of Brentano's theory of consciousness. Despite acknowledging that the latter bears a resemblance to a higher-order theory of consciousness, Rosenthal sometimes draws a parallel between Brentano's theory and his own (ROSENTHAL, 1991, p. 30), while on

[17] For an analysis of this notion of mode of consciousness in Brentano, D. Fisette (2014).

[18] On the idea that mental states are intrinsically conscious, a thesis that Rosenthal attributes to Brentano, see D. Rosenthal (1990, p. 790; 1991, p. 30; 1993, p. 212-213; 1997, p. 30); see also D. Rosenthal (2009, p. 7, 10; 2004, p. 30-31; 2005, p. 179-180, 184).

[19] We can immediately leave aside the objection regarding the individuation of mental states, which Rosenthal (1993, p. 211 *sq.*) addresses indirectly to Brentano, to the extent that it assumes an interpretation Brentano's theory that is in line with that suggested by Kriegel. Such a problem supposes that there is indeed only one (representational) state whose consciousness is an intrinsic property, and the question that Rosenthal asks, and rightly so, is how in these circumstances can mental states be individuated by means of attitudes, such as for example the assertive attitude by which Rosenthal characterizes higher-order thoughts and which differ from non-assertive attitudes such as desire or doubt (ROSENTHAL, 2005, p. 184, p. 180). According to Rosenthal (1993, p. 212-213), a one-level account of consciousness such as Kriegel's, where there are within one single state many parts among which one represents the whole to which it belongs, the criterion of individuation represents a problem for cases of non-assertive attitudes such as desire or doubt: "Suppose the higher-order thought is about a suspicion or doubt; that state will perforce have a mental attitude distinct from any higher-order thought, since higher-order thoughts will invariably have the mental attitude corresponding to an assertion". (ROSENTHAL, 1993, p. 212-213); Kriegel (2003b, p. 487 sq.) responds to Rosenthal's objection with the help of an argument, which rests entirely on Searle's notion of direction of it, which I will not discuss here.

other occasions he associates it with self-representational theories of consciousness (ROSENTHAL, 2009, p. 10) or even with the HOP model of consciousness.[20] The connection with self-representational theories, and more particularly with the version recently upheld by U. Kriegel, seems all the more plausible given that the latter explicitly appeals to Brentano and even characterizes his own theory of consciousness as neo-Brentanian. Moreover, it seems that the thesis that mental states are intrinsically conscious, which Rosenthal attributes to Brentano, rests on an interpretation that is in line with that of Kriegel[21]. Given the impossibility of exposing here in detail the ins and outs of Kriegel's theory, I will simply address here the aspects which enable us to establish a connection with Brentano's theory of consciousness and what justifies, to a certain extent, Kriegel's neo-Brentanianism.

Let us begin by distinguishing Rosenthal's theory from that of Kriegel with the help of the following two definitions, the first corresponding to the HOT theory, and the second to Kriegel's self-representational theory:

1. A mental state *M* of a subject *x* at a given time *t* is conscious iff *x* has a state *M** in such a way that $M \neq M^*$, and *M** represents the occurrence of *M*.
2. A mental state *M* of a subject *x* at a given time *t* is conscious iff *M* represents its own occurrence.

One immediately notices that the main difference between these two theories lies in that the first postulates two numerically distinct mental states, a postulate which the second theory rejects. *M* refers to a single mental state which is nevertheless characterized, as in the higher-order theories, by two distinct contents, the first being the first-order representation, such as the hearing of a sound, while the second is the higher-order content, which corresponds to the consciousness of this representation and, in the present case, the fact of being conscious, oneself, of hearing a sound. In Kriegel's

[20] In connection to the HOP model, Rosenthal (2004, p. 34; 2005, p. 179-180) has made the point that the importance given to inner perception in Brentano's theory of consciousness, like most of the examples taken from visual and auditory perception, seems to indicate that this theory of perception brings it perhaps closer to HOP (higher-order perception) theory than to HOT (higher-order thought) theory. However, this is not the case because inner perception is clearly distinguished from observation or introspection on the basis of its non-reflexive character as Brentano clearly indicates in a text about Thomas Reid's philosophy, whereby he associates observation with reflexive consciousness and inner perception with non-reflexive consciousness. (BRENTANO, 1975, p. 2). This distinction is at the heart of his criticism of introspection in *Psychology* in which he maintains that the accompanying consciousness does not consist in a second-order reflexive act and that the idea of self-observation directed at mental states such as anger, for example, is simply countersensical (*Psychology*, p. 99; see also p. 22). Brentano conceives of inner perception in terms of judgment (*Psychology*, p. 109-110), that is, of *Wahr-nehmung* in the literal sense of the word: as a positive or negative stance (*Stellungsnahme*) taken towards the object of judgment.

[21] On the similarities between Brentano's theory and that of Kriegel, D. Rosenthal (2004, p. 30-31; 2009, p. 10).

theory, however, these two contents are carried by one and the same vehicle which exhibits a particular structure insofar as it consists in a mental state that represents its own occurrence. Hence the thesis that consciousness is in this sense an intrinsic property of mental states.

What makes this theory of consciousness neo-Brentanian in nature is that Kriegel identifies "self-representational consciousness" with the thesis that a mental state is conscious if, and only if, this state is at the same time about itself.[22] Kriegel's interpretation of Brentano's theory of consciousness is consistent with intentionalism insofar as he presupposes not only that intentionality is the single feature of mental phenomena for Brentano, but also that consciousness consists in nothing more than this self-referential structure, or *self-directed intentionality*, by means of which he characterizes mental states. A mental state is therefore conscious if, and only if, it represents its own occurrence. It is in light of this view that Kriegel interprets Brentano's theory of primary and secondary objects, mainly in connection with the following passage from *Psychology* [23]:

> Every mental act is conscious; it includes within it a consciousness of itself. Therefore, every mental act, no matter how simple, has a double object, a primary and a secondary object. The simplest act, for example the act of hearing, has as its primary object the sound, and for its secondary object, itself, the mental phenomenon in which the sound is heard. (*Psychology*, p. 119; *Schriften* I, p. 174).

Considered in itself, this passage seems to corroborate the thesis which Kriegel attributes to Brentano in a recent text ("Brentano's Most Striking Thesis"), and which asserts "that conscious states are conscious *in virtue of* self-representing (and to that extent that self-representation is the essence of consciousness)." (KRIEGEL, 2013, p. 24). Thus, Kriegel supposes that the concomitant consciousness, which in principle accompanies all mental states, is itself a representation and that accordingly the secondary object consists in nothing other than the representation referring to itself as an object. The status of this accompanying consciousness remains admittedly problematic in *Psychology* and we will see that Brentano overcomes some of these problems in his lectures and later writings. However, it is clear that the inner consciousness that accompanies the representation of the secondary object is not itself a representation, but rather a (existential) judgment whose function within

[22] According to Kriegel (2003b, p. 479-480), a neo-Brentanian theory of consciousness is based on the following three theses: the No-Coextension Thesis ("all, but *not only*, conscious states are mental states"); the Physicalist Thesis ("all conscious states *are* physical states"); the Self-Representation Thesis ("all and only conscious states are self-representational states"). Kriegel acknowledges, however, that only the self-representation thesis may be attributed to Brentano!.

[23] This represents the only passage from Brentano's *Psychology* on which Kriegel's interpretation is based and to which he refers on several occasions. (U. KRIEGEL, 2009 p. 14; 2003b, p. 480; 2004, p. 175).

Brentano's theory is similar to the one performed by higher-order thoughts within Rosenthal's theory. On the other hand, the idea of a self-referential structure of intentionality as well as the thesis that consciousness may be reduced to such a self-representational property of mental states is not corroborated by any of Brentano's writings.

Brentano's two theses on consciousness

Let us now return to the point emphasized by Rosenthal (2009, p. 2) that the originality of Brentano's theory with respect to the Cartesian tradition resides in the thesis that all mental states are conscious[24]. This thesis seems to be one of the two general theses formulated by Brentano at the beginning of the second chapter of Book II of *Psychology* (§2):

1. Every mental phenomenon is a consciousness (*Bewußtsein*)
2. Every mental phenomenon is conscious (*bewußt*)

The first thesis refers to the notion of consciousness in its transitive sense, that is, to *consciousness of* something, and thus to intentional consciousness. We may reformulate this thesis as follows:

1b. Every mental phenomenon is consciousness of something.

As a first approximation, the notion of consciousness as expressed in the second thesis is used in an intransitive sense as monadic predicate that refers to an intrinsic and non-relational property of mental states (the fact, for example, that a state like a pain is conscious or unconscious). But this interpretation stands in contradiction with the first thesis since consciousness cannot be at the same time transitive, as in the first thesis, and intransitive as the second suggests. Another interpretation inspired by Brentano's use of the notion of unconscious in *Psychology* (*Psychology*, p. 79; *Schriften* I, p. 120) rests on the distinction established between the passive and the active senses of this notion. The notion of consciousness suggested by the second thesis is comparable to the meaning that Brentano attributes to the notion of unconscious, which he uses in a passive sense, that is, "unconscious" as referring to a thing of which we are (not) conscious", thereby refusing to acknowledge the notion of "unconscious" in an active sense (*Psychology*, p. 79; *Schriften* I, p. 120). In its passive sense, consciousness would therefore refer to the mental phenomenon of which we are conscious or, as Brentano indicates, as an "object of consciousness". Using Brentano's example, we would say that in hearing a

[24] Rosenthal specifies later on in this article: "it was rare until Brentano's time to describe mental states as conscious at all. Even though Descartes and Locke were plainly writing about the property we describe as a state's being conscious, they did not say that our mental states are all conscious, but rather that we are conscious of all our mental states." (ROSENTHAL, 2009, p. 4).

sound, the mental phenomenon of hearing a sound is, in its active sense, about the sound, whereas the act of hearing, in its passive sense, is the object of consciousness insofar as the agent is conscious of being in such a state. We can thus reformulate the second thesis in light of this interpretation that appeals to the distinction between the passive and the active senses of the notion of consciousness:

2b. Every mental phenomenon is an object of consciousness

This formulation fits well with the theory of primary and secondary objects through which Brentano articulates his two theses on consciousness. According to this theory, every mental phenomenon refers at the same time to a primary object (a sound that is heard) and to itself as a "secondary object" (the hearing of the sound). It is to this second thesis to which Brentano devotes the major part of the discussion of consciousness in Book II of *Psychology* and it is on the basis of this thesis that he opposes from the beginning the hypothesis of the existence of unconscious mental states. (*Psychology*, p. 79; *Schriften* I, p. 119).

Now the question remains as to how consciousness can simultaneously stand in relation both to a physical phenomenon (Thesis I) and to itself as an object (Thesis II). Brentano's answer lies in the Aristotelian distinction between the *in recto* and *in obliquo* modes of relation, as the following passage of *Psychology* seems to suggest:

> We can say that the sound is the primary object of the act of hearing, and that the act of hearing itself is the secondary object. Temporally they both occur at the same time, but in the nature of the case, the sound is prior. […] The act of hearing appears to be directed toward sound (*dem Ton zugewandt*) in the most proper sense of the term, and because of this it seems to apprehend itself incidentally (*nebenbei*) and as something additional (*als Zugabe*). (*Psychology*, p. 98; *Schriften* I, p. 146).[25]

As Rosenthal has rightly noted, the difficulty lies in how to interpret this Aristotelian doctrine which plays a central role in Brentano's theory. Indeed, the question is how to understand the *en parergo* relation that consciousness bears to itself as a secondary object[26]. The phrasing of this

[25] This should be compared to the following passage taken from Brentano's lectures on descriptive psychology: "Every consciousness, upon whatever object it is primarily directed, is concomitantly directed upon itself (*geht nebenher auf sich selbst*). In the presenting of the colour hence simultaneously we have a presenting of this presenting. Aristotle already [emphasizes] that the psychical phenomenon contains the consciousness of itself." (BRENTANO, 1982, p. 25).

[26] This is confirmed by a passage from *Psychology* where Brentano identifies his position with that of Aristotle in the *Metaphysics*: "Thus in the twelfth book of the *Metaphysics,* he says, 'Knowledge, sensation, opinion and reflection seem always to relate to something else, but only incidentally to themselves.' Here it is apparent that his conception agrees entirely with our own and he undoubtedly had this conception in mind when he wrote the above quoted passage in which he rejected the infinite complication of mental activity as an unjustified inference". (*Psychology*, p. 102).

passage may be a source of confusion given that the terms "*nebenbei*" (incidentally) and, especially, "*Zugabe*" (something additional) suggest that the accompanying consciousness of the representation of the sound is something extrinsic to the hearing or is to be thought of as a simple additive like the cream or the sugar that one might add to coffee. This further suggests that consciousness would therefore be imposed from without, as higher-order theories maintain, in the sense that the content of the higher-order state would make the target state conscious. But this interpretation is not consistent with Brentano's second thesis on consciousness, which maintains that all mental states are conscious.

There are many ways to understand the dual relationship of consciousness to its primary and secondary objects such as, for example, the distinction between focal and peripheral awareness (or A. Gurwitsch's notion of marginal consciousness or W. James' notion of fringe), which, as we have seen, is used by Rosenthal. We generally refer to this distinction in order to explain the difference between, on the one hand, the attentive and deliberately focused consciousness of things and, on the other hand, the pre-reflexive, non-attentive and immediate consciousness or perception of things. In such a case, the *in recto* consciousness of a primary object would correspond to the focal awareness of a sound while the *in obliquo* consciousness that accompanies the hearing of the sound would correspond to the peripheral awareness of that perception. But this interpretation also entails a number of problems as we will later see.

Brentano and the infinite regress problem

Let us now consider the infinite regress problem that Rosenthal (2005, p. 184) ascribes to Brentano's theory. In Book II of *Psychology*, Brentano examines several objections raised against his own theory, particularly what is known since Aristotle as the threat of infinite regress. This objection is discussed by Brentano in connection with the hypothesis of unconscious mental states as well as with the duplication problem, which refers to the idea that in any mental state a physical phenomenon would have to be represented twice (once in the representation of the sound and once again in the hearing of the sound, that is, the representation of the representation of the sound).[27] The threat of infinite regress is, in fact, the fourth objection addressed by Brentano in §7 (p. 93 ff.)

[27] Rosenthal raises the problem of duplication in Brentano and Aristotle in the following way: "As Brentano puts it, we must choose whether to individuate propositional mental states (presentations) in terms of their (propositional) object or the mental act of the presentation. Brentano credits Aristotle with the idea. Aristotle's actual argument, which Brentano adapts, is that if the sense by which we see that we see is not sight, then the sense of sight and the other sense would both have colour as their proper object, and distinct senses cannot share the same proper object". (ROSENTHAL, 1993, p. 222).

given that his second thesis on consciousness seems to involve such a problem. For if we deny that the representation that accompanies the hearing of the sound is unconscious, as most higher-order theories of consciousness maintain, it would seem that one must therefore necessarily postulate an infinite number of mental states. Brentano's answer consists in denying one of the premises shared by both objections, namely that the concomitant consciousness that accompanies the representation of the sound is numerically distinct from such a representation. Thus, Brentano attempts to demonstrate that both belong to one and the same mental act.

The threat of infinite regress clearly formulated by Brentano (*Psychology*, p.93-94) can be rendered in the following way:

1. Every mental phenomenon is about an object (the hearing of the sound) (Thesis I).
2. Every mental phenomenon is itself the object of an accompanying consciousness (the representation of the hearing of the sound) (Thesis II).
3. The representation that accompanies the initial mental state is numerically distinct from the targeted mental state.
4. If, however, the representation must also be conscious (Thesis II), and the representation that makes it conscious must in turn be conscious, the series is, therefore, infinite.
5. Therefore, either the representation of the initial state is unconscious (and thesis II is, then, false) or there must be an infinite number of mental acts.

The problem lies precisely in the third premise. It posits that the concomitant consciousness, which accompanies the initial representation, is a numerically distinct mental act from the initial mental act to which it refers as an object. Brentano argues that the representation of the sound and the representation of the representation of the sound are one and the same mental act, which is about two different objects, a primary object and a secondary object. From this perspective, the distinction between a lower-order and a higher-order act consists ultimately only in a simple conceptual abstraction:

> The presentation of the sound and the presentation of the presentation of the sound form a single mental phenomenon; it is only by considering it in its relation to two different objects, one of which is a physical phenomenon and the other a mental phenomenon, that we divide it conceptually into two presentations. In the same mental phenomenon in which the sound is present to our minds we simultaneously apprehend the mental phenomenon itself. (*Psychology*, p. 98; *Schriften* I, p. 146).

In other words, there are not two numerically distinct entities, but rather two *abstracta* which belong to one and the same thing, such as, for example,

in the form and the size of a circle or likewise the velocity and the direction of motion[28].

The second assumption, which is challenged by Brentano in his response to this objection, rests on the idea that the concomitant consciousness takes as an object – which refers here to the secondary object – the initial representation as such, that is, the representation of the primary object. This is similar to Rosenthal's theory according to which a higher-order thought can only take as an object the initial or lower-order state.[29] In contrast to Rosenthal, Brentano maintains, however, that the secondary object of the concomitant consciousness consists in the whole mental act, which is comprised of both the represented sound and itself:

> These results show that the consciousness of the presentation of the sound clearly occurs together with the consciousness of this consciousness, for the consciousness which accompanies the presentation of the sound is a consciousness not so much of this presentation as of the whole mental act in which the sound is presented, and in which the consciousness itself exists concomitantly. Apart from the fact that it presents the physical phenomenon of sound, the mental act of hearing becomes at the same time its own object and content, taken as a whole. (*Psychology*, p. 100; *Schriften* I, p. 148).

A review of the objections raised against the second thesis shows, on the one hand, that there is not and cannot be any unconscious representation in the sphere of our experience (*Psychology*, p. 81; *Schriften* I, p. 122) and that, on the other, the threat of infinite regress cannot be considered as an argument against Brentano's theory because the series of mental acts ultimately ends with the second term, that is, with the consciousness of the whole mental act. (*Psychology*, p. 100; *Schriften* I, p. 148).

Three options regarding the Interpretation of "one and the same act"

The question remains now of determining what Brentano means by a "whole mental act" or by the expression "one and the same act" on which rests

[28] As Brentano explains in a fragment published in *Religion und Philosophie*: "Es ist *ein* Akt, den wir nur begrifflich zerlegen, indem wir ihn einerseits denken, insofern er das Farbige, andererseits insofern er das Farbiges-Sehende zum Objekt hat, ähnlich wie wir an einem Kreis Gestalt und Größe oder an einer Bewegung Richtung und Geschwindigkeit unterscheiden". (BRENTANO, 1954, p. 191).

[29] In an appendix to the classification of 1911, Brentano duly insists that if this were the case, the threat of infinite regress would still hold: "As I have already emphasized in my *Psychology from an Empirical Standpoint*, however, for the secondary object of mental activity one does not have to think of any particular one of these references, as for example the reference to the primary object. It is easy to see that this would lead to an infinite regress, for there would have to be a third reference, which would have the secondary reference as object, a fourth, which would have the additional third one as object, and so on". (*Psychology*, p. 215; *Schriften* I, p. 385).

part of his solution to the infinite regress problem, and which is also a presupposition of the doctrine of *in recto* and *in obliquo* consciousness. That is the third problem which Rosenthal associates with Brentano's intrinsicalism, expressed as follows:

> How could we ever show, in a non-question-begging way, that a higher-order thought is part of the mental state it is about, rather than that the two are just distinct, concurrent states? (ROSENTHAL, 1993, p. 212-213).

Providing an answer to this question requires that we first consider the three main options to which the various higher-order theories of consciousness appeal in order to account for the relationship between the representation of the primary object and the representation of the secondary object.

Suppose *M*, the representation of the primary object, and *M**, the representation of the representation or, in other words, the representation of the secondary object. The first version of the account simply consists in identifying *M* with *M**:

1. For any mental state *M* of a subject *S*, there is necessarily a mental state *M** such that *S* is in a state *M**, where *M** represents *M*, and *M** = *M.*

This view has been upheld by Kriegel (2003), but, as of recently, he has endorsed the third option described below (KRIEGEL, 2009, p. 228)[30].

The second option, which is upheld by most higher-order theories of consciousness that subscribe to what van Gulick (2006) has termed the *distinctness assumption*, that is, the assumption that there is a numerical distinction between lower-level and higher-level states, can be characterized in the following way:

2. For any mental state *M* of a subject *S*, there is a mental state *M** such that *S* is in the state *M**, where *M** ≠ *M*.

This position represents views such as Rosenthal's higher-order thought theory, where *M* and *M** are two numerically distinct states. The essential difference between these first two options is that, according to the second view, consciousness is a relational and extrinsic property conferred on the initial state from without by, for example, a higher-order thought whereas, according to the first view, consciousness is an intrinsic property of mental states.

Brentano rejects the second view as indicated by his response to the infinite regress objection, which consists in rejecting the assumption that the representation of the primary object and the representation of the secondary

[30] M. Textor rightly criticizes the various interpretations which identify Brentano's theory with the identity thesis. (TEXTOR, 2006, p. 421-424).

object are numerically distinct. But Brentano also dismisses the first view, as shown by his criticism of phenomenalism[31] and by the following passage of *Psychology* in which he maintains that part of the whole, a "divisive" [32], cannot be identical to another part:

> A divisive never stands in a relation of real identity with another which has been distinguished from it, for if it did it would not be another divisive but the same one. But they do both belong to *one* real entity. (*Psychology*, p. 124-125; *Schriften* I, p. 180-181).

This passage suggests, moreover, that Brentano considers another option, the mereological option, in that he conceives of the representation of the primary object and the concomitant representation of the secondary object as divisives of the same whole (or of the whole mental act).

Hence the third option recently suggested by van Gulick (2006) and Kriegel (2009, p. 228), which postulates a mereological relationship between the primary objects and the secondary objects. Suppose the following three elements:

M^* = Representation of the primary object
M^{**} = Representation of the secondary object
M = The whole (or complex) unifying M^* and M^{**}

3. For any mental state M of a subject S, M is conscious iff there is a M^* and a M^{**}, such that (i) M^* is a part of M, (ii) M^{**} is a part of M, and (iii) M is a whole which M^* and M^{**} are parts of.

According to this view, the consciousness of the primary object and the consciousness of the secondary object are metaphysical parts or, in Brentano's words, divisives that belong to one and the same phenomenon, that is, one and the same reality. This is the view upheld by Brentano in virtue of the principle of the unity of consciousness, to which we will now turn.

Unity of Consciousness

The theory of primary and secondary objects raises what I will here refer to as the complexity problem, that is, the problem of unifying within inner

[31] The phenomenalist hypothesis, which Brentano attributes to A. Bain and W. James, simply consists in identifying the primary objects with the secondary objects as it "assumes that the act of hearing and its object are one and the same phenomenon, insofar as the former is thought to be directed upon itself as its own object. Then either 'sound' and 'hearing' would be merely two names for one and the same phenomenon". (*Psychology*, p. 94; *Schriften* I, p. 140-141).

[32] Brentano justifies the use of the neologism "divisive" as follows: "Naturally, just as we can use one term to cover a number of things taken together, we can also consider each part of a thing as something in itself and call it by its own name. But just as in the first case the object to which the term is applied is not a thing, but a mere collective, the object will not be a thing in this case either. So, for want of a commonly used unequivocal term (since the term 'part' is also applied to real things when they are in collectives) we shall call this a divisive". (*Psychology*, p. 121; *Schriften* I, p. 176).

consciousness the entire complex of elements involved in the constitution of our mental life.[33] Brentano invokes the principle of the unity of consciousness precisely in order to address this problem. The first question raised by Brentano is whether the multiplicity of these elements forms a whole or, rather, a collective (*Kollektiv*), which he defines in the following way. A collective is a multiplicity of parts grouped under the same point of view and each of these parts is an independent thing (BRENTANO, 1954, p. 225).

In contrast to a simple aggregate, a collective such as, for example, a company of soldiers or the trees of a forest may be apprehended from the point of view of a unity and represents in itself a homogeneous totality as Brentano maintains above. However, in contrast to the whole, the parts or, more precisely, the pieces (*Stücke*) maintain their independence in their relationship to the collective, to which they belong as their existence does not depend upon their participation to this whole. Conversely, the collective is neither dependent on the existence of its parts or on the relations between its parts since one can take away a tree or modify the relations between the trees and still talk of a collective.

Such is, however, not the case for wholes such as, for example, a melody whose parts are moments, or what Brentano refers to in *Psychology* as divisives. In contrast to the parts of a collective, divisives stand in a relation of dependence to the whole. In the case of a melody, one may, of course, change the notes of a melody when played in another key, but in order for it to be characterized as one and the same melody, the same relations between the notes must obtained, that is, in the present case, the same chords. We may therefore reframe our initial question and ask ourselves whether the multiplicity of states apprehended in inner perception presents itself as a collective or, rather, as a whole:

> […] in the case of more complex (*verwickelten*) mental states, do we have to assume a collective of things, or, does the totality of mental phenomena, in the most complex states just as in the simplest, form *one* thing which we can distinguish divisives as parts? (*Psychology*, p. 121; *Schriften* I, p. 176).

Brentano's answer is that all mental activity constitutes a whole whose mental states are divisives. In this respect, consciousness of the primary object

[33] The following passage of *Psychology* on which Kriegel's interpretation of Brentano's theory rests gives us a sense of what the complexity problem consists in: "Every mental act is conscious; it includes within it a consciousness of itself. Therefore, every mental act, no matter how simple, has a double object, a primary and a secondary object. The simplest act, for example the act of hearing, has as its primary object the sound, and for its secondary object, itself, the mental phenomenon in which the sound is heard. Consciousness of this secondary object is threefold: it involves a presentation of it, a cognition of it and a feeling toward it. Consequently, every mental act, even the simplest has four different aspects under which it may be considered. It may be considered as a presentation of its primary object, as when the act in which we perceive a sound is considered as an act of hearing; however, it may also be considered as a presentation of itself, as a cognition of itself, and as a feeling toward itself. In addition, in these four respects combined, it is the object of its self-presentation, of its self-cognition, and (so to speak) of its self-feeling". (*Psychology*, p. 119; *Schriften* I, p. 173-4).

and consciousness of the secondary object are both metaphysical parts that belong to one and the same phenomenon and reality. Hence the principle of the unity of consciousness through which Brentano attempts to account for the relationship of these elements as a whole to one and the same reality. (*Psychology*, p. 124-125; *Schriften* I, p. 180-1).

This principle is invoked as early as in the first chapter of Book II in order to understand why multiple mental phenomena which are involved in the simplest of mental acts appear in consciousness not as an aggregate consisting of dispersed elements, but, rather, as a unified reality. It is in this context that Brentano refers to his theory of wholes and parts, whereby mental phenomena are conceived as "partial phenomena (*Teilphänomene*) of one single phenomenon in which they are contained, as one single and unified thing" (*Psychology*, p. 74, translation modified; *Schriften* I, p. 114). This principle reveals itself most significantly in the context both of the complexity problem, which stems from the theory of primary and secondary objects, and of the infinite regress problem, which is insoluble unless one supposes that primary objects and secondary objects form a unified indivisible whole. This point is, furthermore, confirmed by Brentano while discussing the issue of the unity of consciousness:

> [...] the totality (*Gesamtheit*) of our mental life, as complex as it may be, always forms a real unity. This is the well-known fact of the unity of consciousness which is generally regarded as one of the most important tenets (*Punkte*) of psychology. (*Psychology*, p. 126 *Schriften* I, p. 182).

Thus, the purpose of this principle is not to do away with complexity in favor of simplicity, but, rather, to guarantee that what is perceived in inner consciousness is, despite this complexity, something that is unified (TEXTOR, 2006).

Mental agent and self-consciousness

One of the fundamental criteria which Rosenthal associates with higher-order theories of consciousness is the principle of transivity which, as we have seen, stipulates that mental states are conscious if, and only if, one is in some way conscious of that state (ROSENTHAL, 2005, p. 4; 2009, p. 7). It has also been noted that in Rosenthal's theory it is the higher-order thought that performs such a function by positing that in being conscious of a given state one is in a way conscious of oneself as being in that state (ROSENTHAL, 2005, p. 6). To use once again our example, a pain cannot be conscious if the subject does not have a higher-order thought about it, such as "I am currently feeling a pain in my stomach". Thus, the principle of transivity presupposes that state consciousness (pain) is dependent upon subjective (transitive) consciousness insofar as, in

addition to having a higher-order thought, the subject must be conscious of being in such a state or of having it. The question at the present time is to determine whether Brentano's theory complies with this transitivity principle.

To answer this question, I will now turn to some of Brentano's posthumously published writings, written after the publication of *Psychology* in 1874. For, in these writings, Brentano reconsiders his initial theory of consciousness in providing substantial revisions to it. Two of these revisions are particularly relevant in the present context: the first refers to the important distinction between implicit consciousness (or awareness in a wider sense) and explicit consciousness (or awareness in a narrow sense) introduced in the Vienna lectures on descriptive psychology; the second modification consists in the notion of the "mentally active agent" (*Psychisch Tätige*), introduced in several fragments collected in *Religion und Philosophie* as well as in the "Appendix to the Classification of Mental Phenomena" of 1911 to solve some of the problems pending in *Psychology*. I am referring, among other things, to the ambiguous status in *Psychology* of the concomitant consciousness that accompanies all mental states and of the substrate, which Brentano also characterizes as a "unified real being." (*Psychology*, p. 120; *Schriften* I, p. 175), that is, as a being whose modes of consciousness, as divisives, consist in its determinations.

As a first approximation, the notion of a "unified real being" refers to the whole mental state, which consists in a "real" unity. In contrast to physical phenomena, individual mental phenomena "are those phenomena which alone possess real existence apart from (*ausser*) intentional existence". (*Psychology*, p. 70, translation modified; *Schriften* I, p. 109). And, as indicated above, the unity of consciousness consists in these partial phenomena (*Teilphänomene*) belonging to this real thing. But the principle of unity of consciousness, as formulated in *Psychology*, provides us with details neither on the nature of the substrate that underlies and unifies as a whole the modes of consciousness, nor on the status of the simultaneous consciousness that accompanies the various elements that make up this unity. It is precisely in this context that the notion of mental agent is introduced. It first attempts to answer the question as to what constitutes the real substrate of the complex mental act as apprehended in inner perception. This is confirmed by Brentano in a number of fragments that make up *Religion und Philosophie*, and most notably in the following passage where Brentano expresses his general thesis in response to what he calls Aristotle's semi-materialism:

> It therefore follows that one and the same agent must ultimately be at the basis of all mental acts, whether sensory or non-sensory, such as they are simultaneously apprehended in inner perception. The unity of consciousness excludes Aristotle's semi-materialism. (BRENTANO, 1954, p. 228, my translation).

Thus, the modes of consciousness do indeed belong to one and the same complex act as suggested by the principle of the unity of consciousness. However, it is not consciousness as such, but rather the mental agent which is the bearer of this whole. All conscious states are mental phenomena that belong to the mental agent in the trivial sense that it is she, and no one else, who performs these mental acts, and it is she who is conscious of her stomach pain or of the pleasure she takes in playing chess or in composing verses. This privileged and private (or first-person) access to her own mental states is incidentally a presupposition on which Brentano's use of inner perception and consciousness rests.

Hence the second problem which deals with the status of the accompanying consciousness and with the second general thesis on consciousness in *Psychology* according to which all mental states are conscious. This thesis may be interpreted in two different ways whether one conceives the predicate "is conscious" as an intrinsic property of mental states, as Rosenthal sometimes suggests in his interpretation of Brentano, or rather as an object of consciousness in the sense that a mental state is always accompanied by a concomitant consciousness. The first interpretation is problematic for the simple reason that a state as such cannot be said to be conscious (or not) unless one supposes, following G. Ryle, the "self-luminous" character of mental states (D. ROSENTHAL, 1986 p. 344; 1990, p. 738). For, as Brentano (1954, p. 226-228) clearly acknowledges, a state requires that a bearer or an agent performs these acts, and this must be accounted for by an explanation of consciousness. On the other hand, the second interpretation also includes its share of problems since it does not explain why standing in relation to a secondary object would simply make one conscious of performing an act whose object is a physical phenomenon. The problem stands out more clearly in relation to the principle of the unity of consciousness (or that of the consciousness of a real unity): how can consciousness be at the same time both consciousness (in an active sense) of this unity and object of consciousness (in a passive sense), that is, consciousness of an occurring consciousness? While discussing the ideas of Thomas Aquinas, Brentano considers this possibility and maintains that the consciousness of an occurring consciousness coincides with the consciousness of the initial representation. It is precisely in this context that Brentano introduces the idea that the consciousness of the consciousness' representation of the sound is in fact nothing other than the consciousness of the whole mental act as it "becomes at the same time its own object and content". (*Psychology*, p. 100; *Schriften* I, p. 148) But this concomitant consciousness of the secondary object understood as the whole mental act does not take into account the fact that this state is conscious apart from stating that we are conscious of it. Thus, these two explanations of the second thesis, which is at the heart of Brentano's analyses of consciousness in Book II of *Psychology*, do not adequately account

for what it is for a mental state to be conscious. This seems to be what Brentano had later realized, and my hypothesis is that by taking in consideration the mental agent, Brentano attempts not only to resolve the problem of the substrate that underlies the various modes of consciousness, but also to provide a more adequate explanation of the second thesis.

Indeed, it would seem, according to this explanation, that a state is conscious only if an agent becomes aware not of this state as such, but rather of himself as being in such a state. Thus, the appeal to the mental agent in this theory of consciousness implies that in performing normally, say, an act of external perception the agent becomes aware not only of the primary object, but also of himself as a perceiving agent (BRENTANO, 1954, p. 226). This is also confirmed by a passage from the 1911 "Appendix to the Classification of Mental Phenomena" in which Brentano maintains that the object of secondary consciousness or internal perception is the mental agent himself as constituting both the relationship to the primary object and the secondary consciousness as a relation to the agent himself:

> As I have already emphasized in my *Psychology from an Empirical Standpoint*, however, for the secondary object of mental activity one does not have to think of any particular one of these references, as for example the reference to the primary object […] The secondary object is not a reference but a mental activity, or, more strictly speaking, the mentally active agent (*sondern die psychische Tätigkeit, genauer gesprochen das psychisch Tätige*), in which the secondary reference is included (*beschlossen ist*) along with the primary one. (*Psychology*, translation modified, p. 215; *Schriften* I, p. 385).

This passage highlights a new mode of consciousness that is absent from *Psychology*, namely, the mode of consciousness *de se*, which refers to the consciousness of an agent as being oneself in this complex state. Using once again the example of the representation of a sound, self-consciousness would be expressed as follows: I am myself in the process of representing or experiencing a sound.[34] This point stands out even more clearly in the case of pain insofar as it is a state of which the agent is necessarily aware from a first-person perspective. The thesis that all mental states are conscious should then be understood, in light of the *de se* mode of consciousness, as an assertion about the implicitness of this self-awareness in all of experience. To use Rosenthal's vocabulary, this consists in saying that Brentano subordinates subjective consciousness to state consciousness and, then, state consciousness to self-consciousness. In this respect, this new version of Brentano's theory of consciousness is not incompatible with Rosenthal's transivity principle.

[34] This should be contrasted with the remarks made by Kriegel (2003, p. 480-1) regarding the distinction within Brentano between *self-representation* and *representation of the self*.

However, Brentano does not support the view that a mental agent could be transitively conscious of something without being intransitively conscious of being in such a state. In other words, Brentano does not maintain that transitive consciousness can be said to be independent of intransitive consciousness. It is in this sense that I interpret the distinction between implicit consciousness (awareness in a wider sense) and explicit consciousness (awareness in a narrow sense). In these lectures, this distinction is closely associated to the central notion of noticing (*Bemerken*).[35] Brentano first applies this notion to the external perception of a primary object and maintains that one can see or hear (implicitly) something that one does not notice (explicitly). This is demonstrated by Brentano in an example, which recalls an argument made by Dretske (1993) against Rosenthal:

> Whoever sees a lark in the blue of the sky does therefore not yet notice it, and hence will just as little notice his seeing of the lark, even though his seeing of the lark is concomitantly experienced [*mitempfinden*] by him. However, were he, at some point, not only to see the lark, but also to notice it, then he would certainly notice simultaneously that he sees it […].To see is different from being clear about what is seen. And thus, the concomitant experience [*mitempfinden*] of the seeing will be different from being clear about this concomitantly experienced seeing. (1995, p. 26).

Brentano supposes that the lark is not the explicit object of the act even though it appears in the subject's visual field, and that the latter is implicitly conscious of it. This amounts to maintaining that a state may be (implicitly) conscious without the subject being (explicitly) conscious of it. D. Armstrong (1997, p. 723) has also made a similar point with reference to the well-known case of the inattentive driver, which is often considered as exemplifying the use of the notion of unconscious in higher-order theories. Brentano would, in contrast to Armstrong, explain that the driver is not unconscious, but, rather, that he has an implicit and peripheral consciousness of his driving. For not only does Brentano reject the existence of unconscious mental states, but he argues, moreover, that the subject can be explicitly conscious of experiencing something (say, a lark) only if she is implicitly conscious of it (BRENTANO, 1995, p. 36). Explicit consciousness, or consciousness in a narrow sense, constitutes an act of noticing (*Bemerken*) conceived by Brentano in these lectures as an explicit perception of what is implicitly contained in consciousness (BRENTANO, 1982, p. 36). The distinction between implicit and explicit consciousness also helps to dispel some of the obscurities at the heart of the initial theory, most notably regarding the status of mental phenomena in

[35] See K. Mulligan's article "Brentano on the Mind" for a complete analysis of these distinctions in Brentano's lecture on psychognosy.

Psychology, brought to light by Husserl in the *Logical Investigations*[36]. For given that physical phenomena are not elements of inner consciousness insofar as the latter is limited to the domain of mental states, the remaining question is whether this class of phenomena consists of contents of sensory experience or of simple stimulations. The reference to the notion of implicit consciousness shows that *qualia* are elements of primary consciousness and that, in contrast to the view held by higher-order theories, qualitative experience constitutes a necessary condition for having higher-order thoughts, which are about this logically prior experience. It is in this sense that primary consciousness is for Brentano an intransitive and implicit (or intrinsic) consciousness, and as Brentano's commentary on Thomas Reid indicates, it is also a pre-reflective consciousness.[37]

Final remarks

Once we consider the changes that Brentano brings to his initial theory of consciousness, it is clear that one may not reduce it to either versions of the higher-order theory of consciousness. For that matter, Rosenthal's critical remarks about Brentano's *Psychology* confirm this point: Brentano's theory of consciousness is not consistent with the principle of transitivity. In other words, it does not recognize the fundamental principle of any higher-order theory of consciousness. For despite the affinities that hold between Brentano and higher-order theories, most notably with respect to the distinction between the various levels in mental states in his classification of mental acts and in spite of the significance of intentionality in his philosophy of mind, Brentano has never upheld any form of intentionalism whatsoever and has never attempted to reduce consciousness to any type of intentional relation. Rather, consciousness represents within Brentano's theory a form of intransitive self-consciousness which is intrinsic to the agent. Thus, if one admits that the premise at the basis of most of the criticisms addressed to Brentano's philosophy of mind implies mainly this representationalist or intentionalist postulate (also known as "Brentano's thesis"), one must therefore conclude that such criticisms miss their mark and do not do justice to Brentano's original contribution to the analysis of consciousness. For our

[36] D. Fisette (2010) for an analysis of the criticism addressed by Husserl at Brentano in the *Logical Investigations*, and A. Werner (1931) on the ambiguous status of the notion of physical phenomena in Brentano's *Psychology*.

[37] My analysis of Brentano's theory of consciousness is similar, in part, to that proposed by J. Brandl in his article "What is Pre-Reflective Self-Awareness? Brentano's Theory of Inner Consciousness Revisited". Brandl criticizes the higher-order theories' interpretation of Brentano's theory, and defends the view that Brentano upheld a pre-reflective theory of consciousness. However, unlike Brandl, I do not believe that this pre-reflective theory of self-consciousness is already present in the two chapters in *Psychology* on inner consciousness.

analysis of Brentano's writings has shown that his theory of consciousness fulfills most of the requirements that motivate such criticisms and address most criticisms directed at higher-order theories of consciousness. Moreover, Brentano's account of the relationship between consciousness and intentionality deserves to be discussed in greater depth than what was possible in the context of this article.

References

ARMSTRONG, D. "What is Consciousness?". In: BLOCK, N.; FLANAGAN, O.; GÜZELDERE, G. (Eds.): *The Nature of Consciousness:* Philosophical Debates. Cambridge: MIT Press, 1997, p. 721-728.

BRANDL, J. "What is Pre-Reflective Self-Awareness? Brentano's Theory of Inner Consciousness Revisited". In: FISETTE, D. and FRÉCHETTE, G. (Eds.). *Themes from Brentano*. Amsterdam: Rodopi, (2013), p. 41-66.

BRENTANO, F. (*Schriften* I). *Sämtliche veröffentlichte Schriften*, v. 1, *Schriften zur Psychologie, Psychologie vom Empirischen Standpunke / Von der Klassifikation der Psychischen Phänomene*, T. Binder and A. Chrudzimski (Eds.). Frankfurt a. M.: Ontos, 2011.

_____. (*Schriften* II). *Sämtliche veröffentlichte Schriften*, v. 2, *Schriften zur Psychologie, Untersuchungen zur Sinnespsychologie*, W. Baumgartner (ed.), Frankfurt a. M. : Ontos, 2009.

_____. (*Schriften* III) *Sämtliche veröffentlichte Schriften*, v. 3, *Schriften zur Ethik und Ästhetik,* Frankfurt a. M.: Ontos, 2011.

_____. (*Psychology*). *Psychology from an Empirical Standpoint*. Transl. by A.C. Rancurello, D.B. Terrell, and L. McAlister, London: Routledge, 1973.

_____. (2003). *Vom Ursprung sittlicher Erkenntnis*, Leipzig: Dunker & Humblot, 1889; *The Origin of the Knowledge of Right and Wrong*. Trans. by R. Chisholm and E. Schneewind. London: Routledge, 1969.

_____. (1995). *Descriptive Psychology*. Transl. and ed. by B. Müller, London: Routledge; Trans. of *Deskriptive Psychologie*, R. Chisholm and W. Baumgartner (Eds.), Hamburg: Meiner, 1982.

_____. (1975). "Was an Reid zu loben. Über die Philosophie von Thomas Reid", Aus dem Nachlaß herausgegeben von R. Chisholm und R. Fabian. *Grazer Philosophische Studien*, v. I, 1975, p. 1-18.

_____. (1954) *Religion und Philosophie*, F. Mayer Hillebrand (Ed.) Bern: Francke, 1954.

_____. (1930). *Wahrheit und Evidenz*, O. Kraus Ed.). Leipzig: Meiner, 1930.

CASTON, V. (2002) "Aristotle on Consciousness", *Mind*, v. 111, p. 751-815.

CHALMERS, D. "Facing Up to the Problem of Consciousness". *Journal of Consciousness Studies*, v. 2, n. 3, p. 200-219, 1995.

CRANE, T. "Intentionalism. *Oxford Handbook to the Philosophy of Mind*. Oxford: Oxford University Press, (2007).

DESCARTES, R. (1964-1965). "Quatrièmes Réponses". *Œuvres de Descartes*, C. Adam and P. Tannery (Ed.). Paris: Vrin [s.d.].

DRETSKE, F. "Conscious Experience". *Mind*, v. 102, 1993, p. 263-283.

FISETTE, D. "Duas teses de Franz Brentano sobre a consciência", *Phainomenon, Revista de Fenomenologia*, 2014.

_____. «Le «cartésianisme» de Franz Brentano et le problème de la conscience », In: ROUX, S. (Dir.). *Le corps et l'esprit: problèmes cartésiens, problèmes contemporains*, Paris, Éditions des archives contemporaines, 2015. p. 163-208.

_____. "Descriptive Psychology and Natural Sciences. Husserl's Early Criticism of Brentano", C. Ierna *et al*. (Eds.). *Edmund Husserl 150 Years:* Philosophy, Phenomenology, Sciences. Berlin: Springer, 2010, p. 135-167.

FISETTE, D.; POIRIER, P. (2000). *Philosophie de l'esprit. État des lieux*, Paris: Vrin.

GENNARO, R. *The Consciousness Paradox:* Consciousness, Concepts, and Higher-Order Thoughts. Cambridge: MIT Press, 2012.

_____. "Between pure self-referentialism and the (Extrinsic) HOT Theory of Consciousness". In: KRIEGEL, U., and WILLIFORD, K. (Eds.). *Self-Representational Approaches to Consciousness*. Cambridge: MIT Press, 2006, p. 45-66.

GULICK, R. VAN. "Mirror, Mirror - Is That All?". In: KRIEGEL, U., & WILLIFORD, K. (Eds.). *Self-Representational Approaches to Consciousness*, Cambridge: MIT Press, 2006, p. 11-39.

_____. "Inward and Upward. Reflection, Introspection, and Self-Awareness", *Philosophical Topics*, v. 28, 2000, p. 275-305.

GÜZELDERE, G. "Is Consciousness the Perception of What Passes in One's Own Mind?". In: BLOCK, N. *et al*. (Eds.). *The Nature of Consciousness: Philosophical Debates*. Cambridge, MIT Press, 1997, p. 789-806.

INGARDEN, R. "Le concept de philosophie chez Franz Brentano", *Archives de philosophie*, v. XXXII, 1969, p. 458-475 and 609-638.

JANZEN, G. *The Reflexive Nature of Consciousness*. Amsterdam: John Benjamins Publishing.

KASTIL, A. *Die Philosophie Franz Brentanos.* Eine Einführung in seine Lehre, Bern: Francke, 1951.

KRIEGEL, U. "Brentano's Most Striking Thesis". In: FISETTE, D., and FRÉCHETTE, G. (Eds.). *Themes from Brentano*. Amsterdam: Rodopi, p. 23-40.

_____. (2009) *Subjective Consciousness:* A Self-Representational Theory. Oxford: Oxford *University* Press, 2009.

_____. "The Functional Role of Consciousness: A Phenomenological Approach". *Phenomenology and the Cognitive Sciences*, v. 3, 2004, p. 171-93.

_____. "*Consciousness as Intransitive Self-Consciousness*: Two Views and an Argument". *Canadian Journal of Philosophy*, v. 33, n. 1, 2003b, p. 103-132.

_____. "Consciousness, Higher-Order Content, and the Individuation of Vehicles". *Synthese*, v. 134, n. 3, 2003a, p. 477-504.

MARTY, A. *Descriptive Psychologie*. M. Antonelli & J. C. Marek (Eds.). Würzburg: Königshausen & Neumann, 2011.

MORAN, D. "Brentano's Thesis". *Proceedings of the Aristotelian Society*, Supplementary Volumes, v. 70, 1996, p. 1-27.

MULLIGAN, K. "Brentano on the Mind". In: D. Jacquette (Ed.). *The Cambridge Companion to Brentano*. Cambridge: Cambridge University Press, 2004, p. 66-97.

ROSENTHAL, D. "Concepts and Definitions of Consciousness." *Methodology and History of Psychology*, v. 4, n. 3, 2009, p. 55-75.

_____. *Consciousness and Mind*. Oxford: Oxford University Press, 2005.

_____. "Varieties of Higher Order Theory". In: GENNARO, R. (Ed.). *Higher-Order Theories* of Consciousness. Amsterdam: John Benjamin Press, 2004, p. 17-44.

_____. "Thinking that one Thinks". In: DAVIES, M. and HUMPHREYS, G.W. (Eds.). *Consciousness*. Oxford: Blackwell, 1993, p. 197-223.

_____. "The Independence of Consciousness and Sensory Quality », *Philosophical Issues*, v. 1, 1991, p. 15-36.

_____. "Two Concepts of Consciousness". *Philosophical Studies*, v. 49, n. 3, 1986, p. 329-359.

ROSENTHAL, D.; LAU, H. "Empirical support for Higher-Order Theories of Conscious Awareness". *Trends in Cognitive Sciences*, v. 15, n. 8, 2011, p. 365-373.

SCHELL, H. *Die einheit des seelebens aus den principien der Aristotelischen philosophie entwickelt*. Freiburg: Scheuble, 1873.

SIEWERT, C. *The Significance of Consciousness*, Princeton: Princeton University Press, 1998.

SMITH, D. W. "Consciousness with reflexive Content". In: D. W. Smith and A. Thomasson (Eds.). *Phenomenology and the Philosophy of Mind*. Oxford: Oxford University Press, 2005, p. 93-114.

TEXTOR, M. "Neither a Bundle nor a Simple: Brentano on the Unity of Consciousness". In: FISETTE, D. and FRÉCHETTE, G. (Eds.). *Themes from Brentano*. Amsterdam: Rodopi, 2013, p. 67-86.

_____. "Brentano (and some Neo-Brentanians) on Inner Consciousness", *Dialectica*, v. 60, 2006, p. 411–431.

THOMASSON, A. "After Brentano: A One-Level Theory of Consciousness". *European Journal of Philosophy*, v. 8, 2000, p. 190-209.

TUGENDHAT, E. *Selbstbewusstsein und Selbstbestimmung*. Frankfurt a. M.: Suhrkamp, 1979.

WERNER, A. *Die psychologisch-erkenntnistheoretischen Grundlagen der Metaphysik Franz Brentanos*, Hildesheim: Borgmeyer, 1931.

ZAHAVI, D. "Back to Brentano?". *Journal of Consciousness Studies*, v. 11, 2004, p. 66-87.

_____. "Brentano and Husserl on Self-Awareness". *Études Phénoménologiques*, v. 14, 1998, p. 127-168.

ARGUMENTOS
Revista de Filosofia

André Leclerc*

Intentionality or consciousness?

ABSTRACT

I discuss mainly three points in Fisette's target paper: 1) Is it true that consciousness is as fundamental – or even more fundamental – as intentionality is in Brentano's philosophy of mind? I shall try to show that intentionality comes first and sheds light on consciousness in Brentano's work of 1874; 2) I question the idea of self-consciousness as something intrinsic to a mental agent and irreducible to intentionality; 3) finally, is it possible to read Brentano as an intentionalist? I think it is, even if many intentionalists today would not accept Brentano's whole conception of the mind.

Keywords: Brentano; Philosophy of mind; Intentionality; Consciousness; Self-consciouness.

RESUMO

Discuto aqui, principalmente, três pontos do artigo-alvo de Fisette, a saber: 1) É verdade que a consciência é tão fundamental - ou mesmo mais fundamental – que a intencionalidade na filosofia da mente de Brentano? Tento mostrar que a intencionalidade vem primeiro e elucida o papel da consciência no trabalho de Brentano de 1874; 2) questiono a idéia de auto-consciência como algo intrínseco ao agente mental e irredutível à intencionalidade; 3) finalmente, é possível ler Brentano como um intencionalista? Acredito que sim, mesmo que muitos intencionalistas hoje não aceitariam inteiramente a concepção de mente de Brentano.

Palavras-chave: Brentano; Filosofia da mente; Intencionalidade; Consciência; Auto-consciência.

* UFC/CNPq - aleclerc@terra.com.br

Brentano's legacy is certainly among the most important and fascinating in contemporary philosophy. But the interpretation of his philosophical psychology is not always a piece of cake. The Devil lives in the ambiguities of some very important passages. Brentano himself was well aware of that, and his immediate followers as well. [1] Just to make things a little more complicated, there are also some important changes in his philosophical doctrines, especially in 1905 when he rejected his former view of content (an ontological thesis called "intentional in-existence" according to which intentional objects have a special ontological status for being immanent to the content of the state). Kotarbinski dubbed the emerging new doctrine "reism". (see KOTARBINSKI, 1976.) At that point, for Brentano, intentional objects are not anymore *immanent* to the intentional state; they transcend the state, and sometimes they exist, sometimes they don't. The new doctrine creates new problems of its own that could only be overcome with a new doctrine of content to be elaborated by Twardowski and Husserl. Be that as it may, the fact that we still have today new debates on Brentano's work, with people like Dan Zahavi, Uriah Kriegel, Tim Crane and Denis Fisette, should not come as a surprise.

Fisette's paper challenges the perception that most philosophers have of Brentano's philosophical psychology. By doing so, he gives us an opportunity to deepen some of our convictions or to revise them. Of course, any such challenge is always welcome. Just mention the name "Brentano" to anyone with some general philosophical knowledge, and the first word you are likely to hear is "intentionality". Usually, the common view does not go much farther than that. The rest of Brentano's complex philosophical psychology is largely unknown or seems irrelevant. Of course, this is not so. Fisette shows that there is much more to Brentano's philosophical psychology than intentionality. The theory of consciousness is certainly a case in point, and the same holds for the theory of the "mental agent" he shortly presents and puts in the forefront at the end of his paper. What Fisette does is not to deny the importance of intentionality in Brentano's philosophical psychology, but to suggest that we should ponder its importance in the light of other equally important principles and ideas. On that score, I totally agree.

In that short paper, I am not much interested in discussing Rosenthal's Higher Order Theory of consciousness and to compare it with Brentano's theory. I think Fisette has shown convincingly that Brentano gets the upper hand. I would like rather to discuss briefly the following issues raised by Fisette's rich interpretation: 1) Fisette presents some reasons showing that consciousness might be more fundamental in Brentano's psychology than intentionality. In foundational terms, I do not believe it is so. Intentionality seems to me more fundamental and still helps to understand "intransitive"

[1] Namely, Twardowski and Husserl.

consciousness; 2) what does it mean to say that intransitive consciousness is an "intrinsic" property of a mental agent, irreducible to any relation of intentionality? And finally a minor point: 3) is there a possible intentionalist reading of Brentano's work?[2]

1) I believe that intentionality in Brentano's psychology remains in the central position: it is the mark of the mental, the main criterion we apply to decide if a phenomenon is mental or not. No "physical phenomenon" has it, he says. (By the way, Brentano is much more convincing in characterizing mental phenomena, his main concern, than physical phenomena; some examples he gives are quite strange: a landscape would be a physical phenomenon, and others, supposedly, take place in imagination.) Intentionality is the foundational concept, not only of Descriptive Psychology or Phenomenology, but also of Psychology and Philosophy of Mind. Intentionality, more than any other characteristic, is the very the essence of mental phenomena, which is not to say that they don't have any other common characteristic. As a matter of fact, they have. Brentano mentions five such characteristics: 1) All mental phenomena contain intentionally an object in themselves to which they are directed (intentional in-existence); 2) All mental phenomena either are presentations or are based on presentations; 3) They are all given by/in inner perception; 4) They all have an effective existence in addition to intentional in-existence; and 5) They are all given as a unity of consciousness. The second characteristic is disjunctive; it separates all the mental phenomena into two classes instead of saying directly what these phenomena are. The third is very important for Brentano's view of consciousness (more on this soon), but it tells something about the way mental phenomena are given or perceived, not about what they are. The fourth says that their existence cannot be put in doubt, while the existence of physical phenomena always can be; again, it does not tell us what mental phenomena are. And the last one tells us that, contrary to physical phenomena that appear separately or do not appear as parts of a single phenomenon, all mental phenomena appear "as in a unity" given in one single perspective of a conscious agent. It is always possible to distinguish abstractly parts in a mental phenomenon, but the parts are never separated when it is given in inner perception.

Brentano's intention was to capture the essence of mental phenomena in order to distinguish them from physical phenomena. After asserting what is

[2] Just a note before we get started: the way we look at "the problem of consciousness" today, especially when we think of the so-called "hard problem," is very different from Brentano's framework. To pose the same problem in Brentanian terms, we should also consider his Genetic Psychology which consider mental phenomena from a third-person point of view, and not only Descriptive Psychology, which describes mental phenomena from the first-person point of view. This is an important limitation and an important point to bear in mind in this whole discussion.

now known as the Brentano's Thesis (that intentional in-existence is the mark of mental phenomena and that no physical phenomena has it), Brentano declares: "We can, therefore, *define* mental phenomena by saying that they are those which contain an object intentionally within themselves." (BRENTANTO, 1874, p. 75, my italics) But anyone of the five characteristics mentioned can serve to "define" mental phenomena in opposition to physical phenomena. However, if intentionality is not the only trait or characteristic common to all mental phenomena, it is the one that defines them better than any other, as he claims explicitly: "That feature which best characterizes mental phenomena is undoubtedly their intentional in-existence," (p. 75) that is, they all have, as their content, something represented to which they are "directed," not necessarily something existent. Using Locke's vocabulary, we could say that intentionality provides the real essence of mental phenomena; the other traits provide only nominal essences.

Brentano's introduction of Intentionality in 1874 puts together two ingredients that create confusions for his future interprets: directedness and intentional in-existence. The directedness, or mental reference to an object, is the more fundamental trait of intentionality. Brentano took three decades to discover the dead ends of the ontological thesis called "intentional in-existence". Around 1905, he criticized Marty and Meinong for their ontological exuberance and gave up the idea of a special ontological status for the things represented in the content of mental phenomena. Many men died in search of the Eldorado. But the Eldorado they imagined has no special ontological status. It simply never existed. But their thoughts were about a golden city and the content of these thoughts could not be specified in the sentence of a public language without mentioning the Eldorado *in modo obliquo*. In languages with declensions, like Latin, the nominative is the case of categorical reference. The other oblique cases suspend the categorical reference. "Plato's beard" refers to a special beard, not to Plato, "Plato" appearing only in the genitive case. The sentence "Sir Walter Raleigh imagines the Eldorado," specifies the content of a mental state ascribed to Walter Raleigh, and the intentional object is the Eldorado, but there is no categorical reference made to the golden city that appears as accusative (an oblique case) of the verb "imagining."

A stomachache (Fisette's example) is a specific kind of pain, and pain is a sensorial experience. Pain is also a paradigm case of conscious *mental* state. But are pains intentional? The stomachache I feel right now is about/of/directed at… what? Many philosophers think that pains are not intentional. John Searle, Louise Antony and Colin McGinn are regularly cited as members of a group that denies Brentano's thesis precisely for that reason. They take as granted or self-evident that pain, for instance, is not about something, is not directed at something, does not contain (or refer to) a represented object. But pain is certainly a mental phenomenon. Therefore, so the argument goes, intentionality

cannot be the mark of the mental and Brentano is wrong. The alternative would be to adopt consciousness as the mark of the mental, understanding consciousness in a "modal" way: something is mental if and only if it is conscious or *capable of* being conscious ("access consciousness" in Block's terminology).

I think, like most intentionalists, that pains, orgasms, and sensorial experiences in general are intentional. When we are seeing, hearing, tasting, touching or smelling, we are tracking properties outside our bodies from non-conceptual contents "about" changes occurring inside our bodies. These changes we feel are intentional. They indicate something. They point at something. Brentano recognizes this point: "One thing certainly has to be admitted; the object to which a feeling refers is not always an external object." "Still they [the feelings] retain a mental inexistence." (BRENTANO, 1874, p. 69). The famous experience of the phantom limb confirms the fact that a sensorial experience, like the attitudes with conceptual content, can be about something that does not exist. Some people feel an itch in a hand they have lost for years. The itch indicates a localisation in a part of the body that does not exist anymore. Intentionality in Brentano characterizes not only attitudes with a conceptual content, but also conscious sensorial experiences. Fisette says at the end of his paper that we should discuss again the relation between intentionality and consciousness in Brentano's work. I agree: it's a nice program and we should do exactly that.

2) Is there anything like "intrinsic" or "intransitive" (KRIEGEL, 2003, p. 103-132)[3] consciousness in Brentano's Psychology? These two adjectives, I think, might be a bit misleading in this context. *Grosso modo*, a property is intrinsic when its instantiation does not depend on anything but the object that instantiates it. To be made of gold, to have a determinate shape, to have a mass of 3 kilograms are intrinsic in this sense, but not to be married, to be a planet, or to be perceiving an orange. In the context of our discussion, I suppose that "intransitive" means not having a "direct object." (I take it for granted that the relevant sense of the word here is the one it has in grammar, not in logic).

All mental phenomena are given in inner perception. And inner consciousness is the consciousness we have of our own mental phenomena. The knowledge we have of our own mental states is a special kind of knowledge that Anscombe once called "knowledge without observation." That knowledge is immediate, infallible, and non-revisable. The whole and unique source of inner consciousness is inner perception. But inner perception clearly has an object. And *having an object*, as we saw, is something that has to do with intentionality.

[3] Here Kriegel introduces the idea. My interpretation coincides with his: in Brentano, consciousness must be analyzed in terms of intentionality.

I am in the kitchen at midday, thinking and writing about some philosophical problem, when suddenly a blackout happens and only then I realize that the buzz of the refrigerator was there all the time. I perceived the difference only when the buzz stopped. Was I conscious of the buzz? I believe the right answer is "no," and I also believe that this is what Brentano would say. To be conscious of something is to have an object. "We have seen that no mental phenomena exists which is not [...] consciousness of an object." (p. 79).

Brentano says that conscious mental states have two objects: a primary object, the object to which the intentional state is directed, and a secondary object, the mental state itself. I think a relevant question in this discussion would be: Is there anything like a conscious mental phenomenon without a primary object? Brentano's answer is clear when he considers the act of hearing:

> A presentation of the sound without a presentation of the act of hearing would not be inconceivable, at least *a priori*, but the presentation of the act of hearing without a presentation of the sound would be an obvious contradiction. The act of hearing appears to be directed toward [the] sound in the most proper sense of the term, and because of this it seems to apprehend itself incidentally and as something additional. (p. 98)[4]

I wasn't conscious of the buzz in the preceding example because it never was a primary object for me (or for anyone of my mental states at that time), but I became conscious of the interruption of the buzz, as we can be conscious of a shadow, a whole, a gap, a silence between two notes, etc. Chisholm, who was good at recycling medieval distinctions, would say that a primary object could be an *ens per alio* (whose identity depends on something else) as well as an *ens per se* (whose identity does not depend on something else).

If I am right in saying that there is no such thing as a mental phenomena without a primary object— and that includes, we have seen, sensorial presentations like stomachache —, the secondary object, the mental phenomenon itself, appears as an object too for inner perception. Why this could not be understood in terms of intentionality? If "having an object" is part of the *definiens* of what we call "intentionality," there wouldn't be nothing strange in doing so. There wouldn't be self-consciousness (or intransitive consciousness) without a consciousness-of. [5] If we understand by "intrinsic" a quality that something can instantiate in isolation, whose instantiation does not depend on anything else, what exactly is intrinsic in Brentano's theory of consciousness? It seems to me that Brentano's descriptive psychology does not really separate intentionality and consciousness. But intentionality comes first in the logical succession of definitions.

[4] The word "the" in the quote is lacking in the translation.

[5] On that score, I agree with Kriegel's interpretation (2003) that speaks of consciousness in terms of "self-directed intentionality." This is mentioned in Fisette's paper.

3) I believe that part of Brentano's thesis is essentially right. Intentionality is the mark of the mental. And like most "intentionalists" today, I believe it is true even of moods and sensorial experiences. Anything we characterize spontaneously as "mental" exhibits the property of "directedness", that is, they are "about" something, or "of" something. Brentano's thesis is logically stronger than that. It is the conjunction of two theses: 1) intentionality is the mark of the mental, *and* 2) physical phenomena don't exhibit such "aboutness".

The intentionalists defend only the first part of the so-called Brentano's thesis, that is, the intentionality is the mark of the mental, that all the mental acts, states and events are intentional, are about something, or directed to objects. (CRANE, 2014, p. 150)[6] Here "directedness" is the key word. The second part of the thesis says that no physical phenomena are intentional, or directed at something other than themselves. A matrusca doll is not *about* the other dolls it contains, anymore than a rope can be about a hanged man. Intentionalists are not committed to that second part of the thesis. Someone could claim consistently that all the mental is intentional, and nonetheless adopts a reductionist view of the mental as something physical. In that case, if "reducing mental properties" means "identifying them with lower-order properties," and given that identity is symmetric, part of the physical could be seen as intentional. However, this sounds bizarre, because only the mental *qua* mental is intentional. A bunch of neurons cannot be described as intentional. Some token-physicalists, like Davidson, would do exactly this: token-token identity means that part of living matter is mental (by symmetry of "="), but insist on conceptual dualism. There are many physical things that seem to be about something else. But they are not "autonomously" about something, so to speak. "Semanticity," the intentionality of linguistic expressions and other public representations (graphics, photographs, maps, etc.) presupposes the existence of agents capable of using them in a relevant way, and that clearly presupposes mentality. The artefacts, in general, have a proper function that can only be defined by mentioning the intentions, needs and desires of potential users. Smoke, footprints, symptoms, and similar examples of what Grice have called "natural meaning" do not seem to qualify as artefacts. They do not have a proper function and they depend on blind causal relations (fire causing smoke, etc.). Finally, George Molnar (2003) had the idea that physical dispositional properties *tend* to cause their manifestations and possess, therefore, a kind of physical intentionality. It brings some interesting advantages in philosophical psychology and ontology to extend intentionality beyond the realm of mentality; especially, it gives us a unifying view relative to the use of signs and of all sorts of artefacts. But for intentionalists, this is not a main concern.

[6] "For holding that all mental phenomena are intentional does not imply that nothing non-mental is."

Nonetheless, and once again, an intentionalist is committed only to the first part of Brentano's thesis: that intentionality is the mark of the mental. So Brentano could be seen, after all, as an intentionalist *plus* a denial of any form of intentionality in the realm of physical phenomena. In that sense, it is even a bit trivial to say that there is room in Brentano's works for an intentionalist interpretation. Fisette seems to disagree with that.[7] Is it so unreasonable to attribute such a view to Brentano himself?

References

BRENTANO, Franz [1874]. *Psychology from Empirical Standpoint.* Ed. by O. Kraus; English edition by L. McAlister. London and New York: Routledge, 1973/1995.

CRANE, Tim. "Intentionalism" [2009]. In: *Aspects of Psychologism.* Ch. 8. Cambridge (MA): Harvard University Press, 2014.

FISETTE, Denis "Franz Brentano and Higher-Order Theories of Consciousness", in this special issue, 2015.

KOTARBINSKI, Tadeuz. "Brentano as Reist". In: L. McAlister (Ed.). *The Philosophy of Brentano*. Atlantic Highlands: Humanity Press, 1976.

KRIEGEL. "Consciousness as Intransitive Self-Consciousness: Two Views and an Argument", in *Canadian Journal of Philosophy,* v. 33, n. 1, March 2003, p. 103-132.

MOLNAR, George. *Powers. A study of metaphysics.* Oxford: O.U.P., 2003.

[7] See the conclusion of Fisette's paper: "… in spite of the significance of intentionality in his philosophy of mind, Brentano has never upheld any form of intentionalism whatsoever and has never attempted to reduce consciousness to any type of intentional relation."

Revista de Filosofia
ARGUMENTOS

Bruno Leclercq*

Comments on Denis Fisette, "Franz Brentano and higher-order theories of consciousness"

ABSTRACT

For the last few years, research on Brentano's psychology has turned to mereology for a theoretical framework which could help to address and solve some major problems, such as the question of the unity of the mind despite its being made up of lots of simultaneous and consecutive mental acts or the question of the unity of each of these mental acts despite of its being made up of several descriptive components. By using Gilbert Null's formalization of Husserl's mereology we take a closer look at some of Brentano's claims as well as at their issues and consequences.

Keywords: Brentano; Philosophy of mind; Mereology; Husserl, High-order theories.

RESUMO

Nos últimos anos, a pesquisa sobre a psicologia de Brentano vem buscando na mereologia uma base teórica que pudesse ajudar a tratar e a resolver grandes problemas, tais como a questão da unidade da mente apesar de ser constituída por muito atos mentais simultâneos e consecutivos ou a questão da unidade de cada um destes atos mentais apesar de serem constituídos de muitos componentes descritivos. Usando a formalização da mereologia de Husserl feita por de Gilbert Null, podemos examinar mais detidamente algumas reinvindições de Brentano assim como seus problemas e consequências.

Palavras-chave: Brentano; Filosofia da mente; Mereologia; Husserl; Teorias de ordem superior.

* Université de Liège / Belgium. E-mail: b.leclercq@ulg.ac.be

For the last few years, research on Brentano's psychology has turned to mereology for a theoretical framework, which could help to address and solve some major problems in the philosophy of mind[1]. These notoriously include the question of the unity of the mind despite its being made up of lots of simultaneous and consecutive mental acts, and also the question of the unity of each of these mental acts despite of its being made up of several descriptive components. The idea of using mereology as an analytical tool for descriptive psychology was suggested by Brentano himself and even developed by some of his early disciples. In this sense, it is not so much a new idea as the rediscovery of an old one.

The problem, however, is that, besides the general idea of a formal theory of relations between wholes and their parts and apart from a few insights on how such relations could be conceived, there is no single and unified theoretical framework which can be counted as "mereology". Although there is a rather standard formal system for *extensionalist* mereology based on first sketches by Leśniewski and Whitehead and then developed by Leonard and Goodman[2], this system obviously cannot be the tool which descriptive psychology requires. This system, which is built to comply with strong nominalist requirements, only takes wholes as mere sums of their parts so that wholes do not really constitute new entities and parts do not depend on the wholes of which they are part. Therefore, if it is to overcome Hume's bundle theory of mind, descriptive psychology obviously needs some stronger notion of a whole and of its relations to its parts.

In his third *Logical Investigation*, Edmund Husserl notoriously made a first attempt to state in a semi-formal way some of the principles on which such a "stronger" mereology could be grounded. Husserl indeed distinguished between two kinds of parts, namely *pieces* (*Stücke*), which can exist separately from each other and from the whole they are part of, and *moments* (*Momente*), which ontologically depend on the whole of which they are part. He then went on to state some relations which hold between wholes and their pieces as well as between pieces of a same whole; between wholes and their moments as well as between moments of the same whole; between pieces and their own pieces; between pieces and their own moments; and so on. As it is based on wholes and parts, such a formal ontology is a mereology, but, as Peter Simons (1982, 1987) shows, it is very different from Leonard and Goodman's extensional mereology as it involves (several kinds of) dependence relations, which clearly are intensional. Both Kit Fine (1995, p. 463-485) and Gilbert Null (2007, p.33-69;

[1] Besides the authors mentioned in Denis Fisette's paper, I would also mention Arnaud Dewalque, "Brentano and the parts of the mental: a mereological approach to phenomenal intentionality" in Kriegel (2013).

[2] A.N. Whitehead (1916, p.423-454); Leśniewski (1916); H. Leonard and N. Goodman (1940, p.45-55); N. Goodman (1951).

2007, p.119-159) have made attempts to formalize Husserl's theory by interpreting but also completing and systematizing Husserl's insights.

Now, although I do not want to claim that formalization is the only way to guide clear and rigorous reasoning, I believe that it could be very useful to look more closely at Husserl's mereology and its systematization by Fine or by Null both in relation to(1) Brentano's own attempts to think about the mind in mereological terms, and (2) contemporary attempts to solve problems in philosophy of mind by using some of Brentano's notions and theses. By using Null's formal system, I have recently expressed some important differences – and disagreements – between several contemporary readings of Brentano's descriptive psychology as well as drawn some important conclusions from these differences (LECLERCQ, 2014). This work included:

- the debate between Higher-Order Theories of consciousness and several "unilevelist" theories of consciousness[3];
- the debate between the standard conception of intentionality as a relation to some immanent object and Sauer or Antonelli's "continuist" conception of intentionality as both a relation to some transcendent object and a correlation between the act and its immanent content[4];
- the debate between those who do and those who do not identify the phenomenal content of the act with its representational content[5].

All these debates concern some part-whole as well as some dependence relations between components of the mind. And all of them thus lend themselves to some mereological analysis. The reason why Husserl's framework seems to be relevant here is that it seems to fit with Brentano's own mereological claims. In much-discussed pages of his *Descriptive psychology* Brentano does indeed distinguish between parts which are really separable (either *mutually* such as in an act of seeing and a simultaneous act of audition or *unilaterally* such as in an act of presentation grounded on an act of judgement) and parts which are only distinctional". Amongst the latter, Brentano distinguishes between those which are *mutually pervading* such as the affirmative quality of a judgment and its being directed to the object "truth", those that are *logically related* such as the acts of perceiving, of seeing and of seeing red, those which are *correlative* such as the act of seeing and what is seen, and those which are *inseparably concomitant* such as the (primary)

[3] S. Shoemaker (1994, p.21-38); D.M. Rosenthal (1986; 1997; 2005); A.L.Thomasson (2000; 2006); U. Kriegel (2003; 2004a; 2004b, 2006; 2009; 2012; 2013).

[4] R. Chisholm (1967); K. Mulligan and B. Smith (1982; 1985); A. Chrudzimski (2001; 2013); W. Sauer (2006); M. Antonelli (2009); G. Fréchette (2011; 2013).

[5] G. Harman (1990) ; T. Crane (1992); Dretske (1995); M. Tye (1995); U. Kriegel (2003; 2011); B. Loar (2003); G. Graham, T. Horgan and J. Tienson (2007; 2009).

direction of the act upon an object and its (secondary) direction upon itself [6] (1995, p.15-27).

In these pages Brentano explicitly states that intentionality and consciousness are distinctional rather than real parts of the mind but also that they are inseparably concomitant. And this is what we have to give an account of in mereological terms. It is not enough to merely state that, contrary to the claims of the Higher Order Theory of consciousness , intentionality and consciousness are distinctional parts of one and the same act; we still need to know which kind of relation they hold to each other. Brentano claims that they are "inseparably concomitant" rather than "mutually pervading", "logically related" or "correlative". What does that mean?

An interesting feature of Null's formalization of Husserl's mereology is that it distinguishes two different notions of ontological dependence, one being stronger than the other. The basic one, which is called "(weak) founding", simply consists in conditional existence, i.e. in the fact that some object is inseparable from another one, i.e. it cannot exist without the other object also existing. And this relation systematically holds between moments or distinctional parts of the same whole. Unlike pieces, which can exist separately from the whole they are pieces of, moments are ontologically dependent on the wholes they are moments of (Definition 6). And, since Husserl's mereology also admits that wholes are ontologically dependent on their parts – i.e. wholes cannot exist and be what they are without being composed by the parts they are made of (Axiom 4) – it can easily be shown by founding transitivity (Axiom 5) that, unlike pieces of the same whole, moments of the same whole depend on each other, i.e. they require each other in order to exist and be what they are.

But there is also a second and stronger notion of ontological dependence, namely "relative dependence", which allows that, among two interdependent parts of a whole, one be "more fundamental" than the other. Let's first take an example which exceeds the bare field of descriptive psychology and instead concerns the psycho-physical relation: a theory of the relations between mind and body could try (1) to distinguish between a mental state and the neurological state which instantiates it; (2) to state that this mental state ontologically depends on this neurological state; (3) to state that, conversely, the existence of this neurological state necessarily implies the existence of this mental state (so that, in this broad sense of "inseparability", ontological dependence goes on both sides) but still (4) to claim that the physical state is ontologically prior to the mental state and grounds it. According to Null, who claims to follow Husserl on this point, this would require that the grounded component be dependent on some discrete part of the grounding component,

[6] See also K. Mulligan and B. Smith (1985, p.627-644); W. Baumgartner (2013); U. Kriegel (forthcoming).

i.e. on a part of the grounding component which does not overlap the grounded component (Definition 3).

Now, whether we consider that this is a good way to deal with psycho-physical relations or not, it could perhaps help us to think about the relations between intentionality and consciousness. Even if these two were moments of one and the same act – rather than two separate acts as Higher Order Theories suppose – it could still be possible that one of these moments be "relatively dependent" on the other one. And, in principle, this dependency could work in either direction. On the one hand, intentionality could be more fundamental and consciousness could (always) "come on top of it". Consciousness would somehow supervene on the intentional act. Or, on the other hand, consciousness could be more fundamental, something like the very basis of the mind, and intentionality would (always) come on top of it. Consciousness in general would be the essential feature of the mind, which intentionality, i.e. "consciousness of…", could specify by directing it towards some specific object in some specific way.

By stating that intentionality and consciousness are "inseparably concomitant", Brentano seems to claim that neither of them is less fundamental than – and "comes on the top" of – the other. Some parts of Brentano's investigations, however, could support other readings.

The whole discussion about whether there are unconscious intentional acts seems to show that consciousness presupposes the intentional act which it makes aware of. And of course this is what led to the Higher-Order Theory of consciousness. But even without taking intentionality and consciousness to be separate mental acts as HOT does, it could be possible to consider that the first of these inseparable components of a single mental act is more fundamental than the second one. Despite being inseparable from intentionality, consciousness would be "incidental" (*nebenbei*) and "additional" (*als Zugabe*) to it.

In contradiction to all this, some pages of Brentano's *Theory of categories* seem to suggest that intentionality comes on top of consciousness. Brentano indeed talks about the mind as a substance and about the thinker or the auditor (i.e. some specific intentional instantiations of the mind) as its accidents. And he explicitly uses mereological terms to give an account of this: since the mind can "survive" the disappearance of the thought or the audition while the thinker or the auditor cannot exist without the mind, mind is said to be part of the thinker and of the auditor, which unilaterally depend on it. The problem, however, is that while Brentano claims that the thinker as a whole is something more than the mind, he also claims that there is no other part which completes the mind to make it a thinker; the accident of the mind which makes it a thinker is nothing, i.e. it is no real thing which could itself be considered as a separate object (1981, p. 115-116). This, as Barry Smith has underlined, makes that part of Brentano's mereology problematic as it violates

the weak supplementation principle in such a way that we can barely see such a theory as being still a mereology, i.e. as considering wholes being made of parts (1994, p.70-73)[7]. Brentano says that mind is not so much "completed" as "modified" by thought to make it a thinker; that thought is less a part of the whole than one of its "modalities".

This either forces us to give up regarding Brentano's theory of substance and accident as a genuine mereology or to reinterpret it as merely saying that the accident is not a *piece* – i.e. an independent part – yet a part of the whole; it is just a *moment*, a distinctional part of the whole. According to Smith, the reason why Brentano did not put things that way is that he started from Aristotle's standpoint which would not even consider that the bare mind and the thinker could both exist at the same time; when one actually exists the other only has potential existence, so that they cannot sustain part-whole relations (SMITH, 1994, p.78-79). If however we consider that the bare mind as a substance is part of the thinker as a whole – as Brentano seems to do – we could consider that the thought as the accident of the mind is another part of the whole, though only a distinctional and not a real part of it. The intentional thought would then not only be dependent on the thinker as a whole but also be less fundamental than and "relatively dependent" on the mind; it would come on top of it.

Such an asymmetry would notably lie in the fact that, even though mind is bound to be intentionally oriented towards some object and is therefore generically dependent on some intentional act – consciousness is bound to be consciousness of something – it is not ontologically dependent on this particular intentional act rather than another, while this particular intentional act seems to be ontologically dependent on this mind rather than generically dependent on some mind. This is how I take Denis Fisette's claim that intentionality not only involves consciousness but *de se* consciousness, i.e. consciousness of being the mental act of some particular mind. In other words, intentional acts are "accidents" of – and ontologically dependent on – particular minds; there is no general thought of the Eiffel Tower which would generically depend on some mind but not ontologically depend on any particular mind; my thought of the Eiffel Tower is not the same thought as Denis Fisette's thought of the Eiffel Tower because it involves some implicit reference to my mind as its bearer.

Now, how can we reconcile this idea that consciousness, taken as some personal mental agency, comes first and is then specified or modalised by particular intentional acts with the idea that consciousness yet presupposes the intentional act which it makes aware of? Does being conscious of thinking of the Eiffel Tower not somehow "come on top of" thinking of the Eiffel Tower? Being conscious in general necessarily implies some thinking but does not

[7] See also Chisholm (1982, p.3-16).

depend on any particular thought. Yet being conscious of thinking of the Eiffel Tower depends on a particular thought (and this is what made HOT plausible).

In order to give an account of the relations between consciousness and intentionality which goes beyond the mere claim that they are distinctional parts of the same act— which I think is what Denis Fisette tries to do in this paper — we probably need to distinguish between consciousness in general, which is generically dependent on some intentional act though not ontologically dependent on any particular one, and consciousness of some particular intentional act, which is ontologically dependent on this particular act. While, according to Brentano, any particular intentional act is "inseparably concomitant" of the consciousness of it (which is a symmetrical relation), it seems to be "relatively dependent" on consciousness in general (which is an asymmetrical relation). And of course, consciousness of a particular act is "logically related" to consciousness in general: being conscious that one sees red is an instantiation of being conscious.

Even though they surely are much more complex than extensional part-whole relations, all these relations between distinctional parts of a mental act seem to be within reach of a richer mereology such as Husserl's system (as it is formalized by Null) which uses two notions of ontological dependence.

References

ANTONELLI, M. "Franz Brentano et l'"inexistence intentionnelle"", *Philosophiques*, v. 36 n. 2, p. 467-487, 2009.

BAUMGARTNER, W. "Franz Brentano's mereology". In: D. Fisette and G. Fréchette (Eds.). *Themes from Brentano*, Rodopi, Amsterdam, 2013.

_____. *The theory of categories*. The Hague, Martinus Nijhoff, 1981.

BRENTANO, F. *Descriptive psychology*, London: Routledge, p.15-27, 1995.

CHISHOLM, R. "Intentionality." In: P. Edwards (Ed.). *Encyclopaedia of Philosophy*, London: MacMillan, 1967.

"BRENTANO on Descriptive Psychology and the Intentional". In: E. Lee and M. Mandelbaum (Eds.). *Phenomenology and Existentialism*. Baltimore: John Hopkins University Press, p. 1-23;

_____. "Brentano's Theory of Substance and Accident", in *Brentano and Meinong Studies*, Amsterdam, Rodopi, 1982.

CRANE, T. "The Nonconceptual Content of Experience". In: *The Contents of Experience*. Cambridge, Cambridge University Press, p.136-157,1992.

DRESKE, F. *Naturalizing the Mind*. Cambridge MIT Press, 1995.

CHRUDZIMSKI, A. *Intentionalitätstheorie beim frühen Brentano*, Kluwer, Dordrecht 2001.

_____. "Brentano and Aristotle on the ontology of intentionality". In: D. Fisette and G. Fréchette (Eds.). *Themes from Brentano*, Rodopi, Amsterdam, 2013.

DEWALQUE, Arnaud "Brentano and the parts of the mental: a mereological approach to phenomenal intentionality" U. Kriegel (Ed.). "Phenomenal Intentionality Past and Present", Special issue of *Phenomenology and the Cognitive Sciences*, online first published in January 2013. (DOI: 10.1007/s11097-012-9293-8), 2013.

FINE, K. "Part-whole". In: B. Smith and D. W. Smith (Eds.). *The Cambridge Companion to Husserl*. Cambridge University Press, 1995, p. 463-485.

FRÉCHETTE, G."Deux concepts d'intentionnalité dans la *Psychologie* de Brentano". *Revue roumaine de philosophie*, v. 55, p. 63-86, 2011.

_____. "Brentano's Thesis (Revisited)". In: D. Fisette and G. Fréchette (Eds.). *Themes from Brentano*: Rodopi, Amsterdam, 2013.

GRAHAM; HORGAN; TIENSON. "Consciousness and Intentionality". In: M. Velmans and S. Schneider (Eds.). *The Blackwell Companion to Consciousness*. Oxford: Blackwell, 2007, p. 468-484.

_____. "Phenomenology, Intentionality, and the Unity of the Mind". In: S. Walter, A. Beckermann, and B. McLaughlin (Eds.). *The Oxford Handbook of Philosophy of Mind*. Oxford: OUP, 2009, p. 512-537.

GOODMAN, N. *The Structure of Appearance*. Cambridge: Harvard University Press, 1951.

HARMAN, G. "The intrinsic quality of experience". *Philosophical perspectives,* v. 4, 1990, p. 31-52 .

KRIEGEL, U. "Brentano's Mereology". In: U. Kriegel (Ed.). *Routledge Handbook of Brentano and Brentano School*, forthcoming. [s.d.].

_____. "Consciouness as Intransitive Self-Consciousness: Two Views and an Argument", *Canadian Journal of Philosophy*, v. 33, 2003, p.103-132.

_____. "*A Functional Role of Consciousness*: a Phenomenological Approach". *Phenomenology and the cognitives Sciences*, v. 3, 2004a, p. 175-176.

_____. "Consciousness and Self-Consciousness". *Monist, v.* 87, 2004b, p.185-209.

_____. "Same-Order Monitoring Theories of Consciousness". In: U. Kriegel and K. Williford (Eds.). *Self-Representational Approaches to Consciousness*, MIT Press, Cambridge (Mass.), 2006.

_____. *Subjective Consciousness:* a Self-Representational Theory of Consciousness. New-York: Oxford University Press, 2009.

_____. *The Sources of Intentionality*. New-York: Oxford University Press, 2012.

_____. "Brentano's Most Striking Thesis: No Representation without Self-Representation". In: D. Fisette and G. Fréchette (Eds.). *Themes from Brentano*. Rodopi, Amsterdam, 2013.

_____. *The Sources of Intentionality*. Oxford: OUP, 2011.

_____. (Ed.). "Phenomenal Intentionality Past and Present". Special issue of *Phenomenology and the Cognitive Sciences*. Online first published in January 2013 (DOI: 10.1007/s11097-012-9293-8).

LOAR, B. "Phenomenal Intentionality as the Basis of Mental Content". In: M. Hahn, and B. Ramberg (Eds.). *Reflections and Replies:* Essays on the Philosophy of Tyler Burge. Cambridge (Mass.): MIT Press, 2003, p. 229-258.

LECLERCQ, B. « Dépendance ontologique et intentionalité : relecture méréologique de quelques débats d'interprétation de Brentano», paper presented at the conference *L'ontologie des relations* held in Concordia University (Montreal) on May 12 2014.

LEŚNIEWSKI, S. *"Foundations of the General Theory of Sets, I".* [Podstawy ogólnej teorii mnogości I]. Moscow: Popławski, 1916.

LEONARD H.; N. N. GOODMAN. "The Calculus of Individuals and its Uses". *The Journal of Symbolic Logic*, v. 5, n. 2, p. 45-55, 1940.

MULLIGAN, K.; SMITH, B. "Franz Brentano on the Ontology of Mind". Review of Franz Brentano's *Deskriptive Psychologie* (Hamburg: Meiner, 1982), *Philosophy and Phenomenological Research*, v. 45, p.627-644, 1985.

NULL, G. "The ontology of intentionality I". *Husserl Studies*, v. 23, p. 33-69, 2007.

_____. "The ontology of intentionality II". *Husserl Studies*, v. 23, p.119-159, 2007.

SHOEMAKER, S. "Self-reference and Self-awareness". *The Journal of Philosophy, v.* 65, 1968, p. 556-79; "Phenomenal Character", *Nous*, v. 28, p. 21-38, 1994.

SIMONS, P. "The formalisation of Husserl's theory of wholes and parts". In: B. Smith (Ed.). *Parts and moments*. Philosophia Verlag, 1982, p.111-159; P. Simons, *Parts: a Study in Ontology*. Oxford: Clarendon Press, 1987.

ROSENTHAL, D. M. "Two concepts of Consciousness". *Philosophical Studies*, v. 49, 1986, p. 329-359; "A theory of Consciousness. In: N. Block, O. Flanagan and G. Güzeldere. *The Nature of Consciousness:* Philosophical Debates, 1997, p. 729-754.

_____. "How many kinds of consciousness?", *Consciousness and Cognition*, 11, p.653-665; *Consciousness and Mind*. New-York: Oxford University Press, 2005.

THOMASSON, A.L. *"After Brentano:* A One Level Theory of Consciousness". *European Journal of Philosophy*, v. 8, p.190-209. 2000.

_____. Self-Awareness and Self Knowledge". *Psyché,* 12, p. 1-15, 2006.

SMITH, B. *Austrian Philosophy:* the Legacy of Franz Brentano. Open court, Chicago, 1994.

SAUER, W. "Die Einheit der Intentionalitätskonzeption bei Brentano". *Grazer philosophische Studien,* v. 73, p. 1-26, 2006.

TYE, M. *Ten Problems of Consciousness:* A Representational Theory of the Phenomenal Mind. Cambridge (Mass.): MIT Press, 1995.

WHITEHEAD, A.N. "La théorie relationniste de l'espace". *Revue de métaphysique et de morale*, v. 23, n. 3, p. 423-454, 1916.

Revista de Filosofia
ARGUMENTOS

Diana I. Pérez*

What is it like to be HOT?

ABSTRACT

In a recent paper, Fisette tries to show that Brentano's theory of consciousness can be considered as a higher order theory of consciousness (HOT), and a better one than Rosenthal's because Brentano —unlike Rosenthal— can answer all the objections traditionally posed to HOT theories, introducing the idea of self-consciousness and the distinction between implicit and explicit consciousness. In this paper, I will first reconstruct Fisette's main points, and then I pose some questions to his version of Brentano´s theory. Finally I add some further reasons to reject higher order theories of consciousness.

Keywords: Philosophy of mind; Brentano; Higher order theory of consciousness; Consciousness.

RESUMO

Em um artigo recente, Fisette tenta mostrar que a teoria da consciência de Brentano pode ser considerada como uma teoria de ordem superior da consciência (HOT), e uma teoria melhor que a de Rosenthal, porque Brentano, diferentemente de Rosenthal, pode responder a todas as objeções tradicionalmente feitas às teorias HOT, ao introduzir a ideia de uma auto-consciência e da distinção entre consciência implícita e explícita. Neste artigo, eu primeiramente reconstruirei os pontos principais de Fisette, e então questionarei a sua versão da teoria de Brentano. Finalmente, proponho algumas razões adicionais para rejeitar teorias de ordem superior da consciência.

Palavras-chave: Filosofia da mente; Brentano; Teoria de ordem superior da consciência; Consciência.

* UBA-CONICET/IIF/SADAF. E-mail: dianazerep@gmail.com

Consciousness is, without any doubt, one of the most puzzling issues in philosophy. It is not surprising that so many people tried to give an account of this astonishing phenomenon. In his paper, Fisette analyses carefully Brentano's theory, and tries to show that Brentano's theory of consciousness can be considered as a version of the higher order theory of consciousness (HOT), and a better one that Rosenthal's. The central idea, in Fisette's words, is to show that: "Brentano subordinates subjective consciousness to state consciousness and then, state consciousness to self –consciousness." (p. 30). And in so doing, Brentano seems to have a theory similar to – but better than – Rosenthal's, because he can answer all the objections traditionally posed to HOT theories. It is in order to answer them that Brentano introduces the idea of self -consciousness and the distinction between implicit and explicit consciousness.

In this paper I will first reconstruct Fisette´s main points, and then I will pose some questions to his version of Brentano´s theory. Finally I will add some further reasons to reject higher order theories of consciousness.

I

One of the main questions posed in Fisette's paper is whether Brentano should be read as a defender of a HOT theory of consciousness or not. HOT theories are reductive theories of phenomenal consciousness, a special kind of functionalist/representationalist theories. A higher order thought theory of consciousness claims that a given state, let us say the pain I am feeling right now, is conscious if and only if it is accompanied by a specific thought about that very pain, a thought that could be expressed as "I am presently feeling pain". The higher order mental state, the thought that makes conscious my pain, is a contentful state whose content involves a relation between the pain and me. This is what Rosenthal calls the "Transitivity principle" according to which the intransitive consciousness of the pain state depends upon the transitivity of the higher order state which is *about* the pain.

According to Rosenthal, Brentano's theory is not a HOT theory because, unlike HOT theorists, he does not understand consciousness as an "extrinsic, transitive and relational property of mental states." (p. 7), but as an intrinsic one. As I said above, according to HOT theories, there are two different (i.e. numerically distinct independently existent) mental states, a lower order state and a higher order one, and it is because the second one is about the first one that the first one becomes conscious. In the standard theory, the second order state is not intransitively conscious unless a third order state takes it as its intentional object. But Brentano, according to Rosenthal, also held that all mental states are conscious, and therefore he had to face the infinite regress

objection: given the fact that a higher order state is needed in order to make conscious each mental state, an infinite number of higher and higher order states are needed in order to make all of them conscious. This is the first problem Fisette poses to Brentano. But there are two more difficulties, which are related to each other. First, the relation between first order and second order mental states should be explained. And second, there is the problem of individuating these states, derived from the fact that it is not as clear as it seems whether Brentano held that first order and second order states were two different mental states (as Rosenthal holds) or just one state, i.e. it is not clear if he claimed that the first order and the second order states should be identified (because in the end, according to Rosenthal, it is not clear what is the connection between them, in Brentano's theory).

According to Fisette, Brentano answers these objections with the thesis of the unity of consciousness. The peculiar way in which Brentano answers to the question about the relation between first and second order states, conceiving the unity of these two states as a single mental act in which both states are "divisives" (p. 24-25) -i.e. constitutive parts of the very same act- is the key to face all of Rosenthal's objections. Because, this explanation of the relation between first order and second order states, avoids the infinite regress and answers at the very same time the question about how many states are there (the individuation problem).

But in order to be properly called a HOT theory of consciousness, Brentano should accept the transitivity principle, and it is not clear whether he accepted it or not. According to Fisette, Brentano's theory can be seen as a HOT theory if we take into consideration two ideas that are presented in his posthumous writings: first the distinction between implicit and explicit consciousness and second the idea of a mentally active agent. The first distinction is explained with a familiar example: the one of the driver who did not pay attention to the road, but who was implicitly conscious of it (although not explicitly: he did not pay attention to how many lights were in the border of the road, so he did not count them, but he was implicitly conscious of them because he did not pass any red light while driving). The second idea –the mentally active agent- is needed in order to give a proper account of the complexity of the conscious mental act: it is the mental agent who is conscious of himself in the process of experiencing X (a sound for example), who becomes conscious of X (the sound) i.e. while he is thinking about (transitively conscious of) his experience, the experience becomes (intransitively) conscious. We can see now why Fisette said what I quoted at the very beginning of this paper: "Brentano subordinates subjective consciousness to state consciousness and then, state consciousness to self –consciousness." (p. 30). With all these pieces at hand the puzzle can be solved: implicit consciousness (first order mental states) are what in the literature are called *qualia*, the elements of primary or pre-reflective

consciousness which are according to this reading of Brentano's view a necessary condition for having higher order thoughts, and hence for having transitive self-conscious mental states.

II

I will not discuss Fisette's historical points about Brentano, neither the interpretation of Brentano he offers. I will pose some problems that I think can be raised against the account of consciousness attributed to Brentano by Fisette.

In the first place, it is not clear to me that the theory attributed to Brentano could be understood as a HOT theory, if –as I understand them- these theories are seen as reductive theories. According to Fisette's reading, Brentano is offering a theory of phenomenal or subjective consciousness (*qualia*, for short); in the beginning of the paper Fisette announce that Brentano was engaged in the project of solving the "hard problem" of consciousness, following Chalmers' words. So, he seems to be accepting the classical distinction between phenomenal vs. psychological consciousness offered by Chalmers 1996. And usually HOT theories of consciousness are considered as reductive materialist theories of phenomenal consciousness (and, in this sense, opposed to other non reductive dualist theories, such as Chalmers' one). But in the end of the article (p. 31) it seems that *qualia* are just necessary conditions for higher order consciousness and hence that higher order thoughts should no be identified with *qualia*, therefore the project was not to give a reductive account of *qualia* after all. What the distracted driver case shows seems to be that there are some implicit, pre-reflexive, phenomenal conscious first order mental states that are not the objects of any thought we actually have. But if, as Fisette says, qualitative experience constitutes only a necessary condition for having higher order thoughts, and they cannot be identified with second order thoughts as reductive theories hold, in what sense phenomenal states are conscious? Are they first order conscious? If the answer is yes, then HOT theories are superfluous, because we already had first order conscious states! Second order states are unnecessary in order to understand first order states as conscious.

In the second place, it is important to keep in mind that the distracted driver case is usually mentioned in the philosophical literature in order to distinguish between phenomenal consciousness (*qualia*) and psychological consciousness (or access consciousness) (BLOCK, 1995; CHALMERS, 1996). The idea is that some states are phenomenally conscious in the sense that they do not have any impact on the rational control of our behavior, they have no consequences in our actions or further thoughts. *Qualia* are just the way in which it feels like to be in a given state. And the *qualia* literature usually aims to show that both kind of states are distinct and can exists independently. But

both of them seem to be conscious, *qualia* are pre-reflexively conscious, or implicitly conscious, while psychological states are explicitly conscious. HOT theories – as the reductive theories they are- deny the existence of conscious first order state which are not constitutive part of second order states. But the cases mentioned seem to point to some first order states, which are pre-reflexively conscious without being psychologically conscious, i.e. without being the subject of any second order thought. If this is so, then it seems that, in the end, Brentano himself in his last writings denied higher order theories of consciousness and favored first order ones, because he accepted the existence of pre-reflexive conscious states of mind. If this is Brentano's view, I would be delighted, I defended elsewhere the idea that second order theories of consciousness are wrong. (PÉREZ, 2008).

In the third place, Rosenthal's HOT theory is not, in my opinion, the best version of HOT theories, because it requires that the first order state is an actual part or subject matter of a second order thought in order to be conscious. And because of that it cannot make room to cases like the distracted driver, where some first order states seem to be phenomenally conscious without being psychologically conscious. But dispositional HOT theories like the one defended by Peter Carruthers (2005), where a first order state is conscious just in case it *can* be the part or subject of a second order thought without being *actually* so, can incorporate those cases without abandoning the HOT theory. May be this is the version of HOT theories that Fisette thinks Brentano could have been defending. If this is so we can have a reductive HOT theory of consciousness and accommodate the distracted driver case, i.e. we can incorporate Brentano's pre-reflexive consciousness.

But as I said above, I do not accept myself HOT theories of consciousness because they are too demanding: they require that the subject can have thoughts, sometimes quite complex, involving some concepts such as "self" (as Brentano seems to demand with his idea of a mentally active agent), the concepts involved in mind reading abilities (in Carruther's version), or psychological concepts like "pain" (as a constitutive part of the thought "I am in pain" which makes conscious my pain state); concepts which does not seem to be available to some creatures which all of us would agree that can have some conscious mental states, creatures such as babies and probably some non human primates. So let me introduce, in the next section, some general worries against HOT, following this line of thought.

III

In this last part of this paper, I would like to address the more basic question about the plausibility of HOT theories of consciousness in general. I

think that HOT theories of consciousness have many flaws that are not solved in Fisette's paper. In Pérez (2008), I objected the arguments given by Carruthers (2005) in order to prefer HOT theories instead of first order theories of consciousness and I still prefer these ones. So, I will try to develop in this last section of this paper the last suggestion I made above in order to reject HOT theories of consciousness.

The idea is simple: we should distinguish between the way in which we think and talk about phenomenal consciousness from phenomenal consciousness itself. In my opinion, the defenders of HOT theories of consciousness confuse both. For example, Carruthers (2005) offers six desiderata for a successful reductive theory of phenomenal consciousness. He says that a theory like this should explain (1) why phenomenally conscious states have a subjective aspect to them; (2) why there should seem to be such a pervasive explanatory gap between all the physical, functional and intentional facts, on the one hand, and the facts of phenomenal consciousness, on the other; (3) why people believe that the properties of their phenomenal experience are *intrinsic*, being non-relationally individuated; (4) why their possessors consider phenomenally conscious experience *ineffable,* (5) *private and* (6) *infallible, not just privileged known.*

Carruthers' strategy is to show that his HOT theory can explain these features, while first order theories cannot. But note that except for (1) all the other desiderata are concerned with *the way in which we conceptualize our experience*, that is, the way in which people think or know their experience, not with the experience itself. And in my opinion, the way in which we think or talk about our conscious states trivially presupposes that we can have thought about our conscious states involving concepts such as "self", "experience", "feel", "see", "pain", and so on. But it is not obvious that we should possess those concepts in order to be in the conscious mental state itself. Returning to the distracted driver case, we can say, for example, that the conscious experiences he had were not conceptualized and that is the reason why they did not enter into the rational decision making system, or were not stored in memory. But they were conscious in the sense that, if the co-driver asked the driver during the trip, in the appropriate moment, if he was seeing the red light he would have answered yes. But babies and non-human primates who do not possess complex concepts are able to have these conscious states without being able to conceptualize or report them. They cannot write a book about phenomenal consciousness but this is not a reason to say that they are not conscious. So we should not require the conceptual complexities that HOT theories require in order to explain phenomenal consciousness. Taking these considerations into account I think we should tip the balance towards first order theories of consciousness. HOT theories show the sin that many other philosophical theories show: they take the typical adult human being as the

paradigm in order to develop a philosophical theory; and as a consequence they cannot accommodate non typical examples.

References

BLOCK, N. "On a Confusion about a Function of Consciousness," *Behavioral and Brain Sciences*, v. 18, n. 2, p. 252-253, 1995.

CARRUTHERS, P. *Consciousness:* essays from higher order perspective. OUP, 2005.

CHALMERS, D. *The Conscious Mind.* OUP, 1996.

FISETTE, D. *Published in this issue*. "Franz Brentano and higher order theories of consciousness", 2015.

PÉREZ, D. "Why should our mind-reading abilities be involved in the explanation of phenomenal consciousness? *Análisis Filosófico*, v. XXVIII, n. 1, 2008.

ARGUMENTOS
Revista de Filosofia

Guillaume Fréchette*

Brentano's soul and the unity of consciousness

ABSTRACT

In the following paper, I discuss Fisette's reconstruction of Brentano's view, according to which Brentano's conception of consciousness and of its unity is based on the presupposition that consciousness has a bearer, i.e. the soul. First, I identify Fisette's real target (sect.1) and challenge his conception of the mental agent as central to Brentano's account (sect. 2 and 3). In section 4, I formulate some doubts about the sources used by Fisette, and, in section 5, I propose another reading of the relation between the unity of consciousness and the mental agent in the late Brentano.

Keywords: Philosophy of mind; Brentano; Soul; Consciousness.

RESUMO

No seguinte artigo, discuto a resconstrução de Fisette da visão de Brentano, de acordo com a qual a concepção da consciência de Brentano e a sua unidade é baseada na pressuposição de que a consciência tenha um portador, i.e., uma alma. Primeiramente, identifico o alvo real de Fisette (sec. 1) e desafio a sua concepção de agente mental como central para a teoria de Brentano (sec. 2 e 3). Na seção 4, formulo algumas dúvidas sobre as fontes usadas por Fisette, e na seção 5, proponho outra leitura da relação entre a unidade da consciência e o agente mental no Brentano tardio.

Palavras-chave: Filosofia da mente; Brentano; Alma; Consciência.

* University of Salzburg. E-mail: guillaume.frechette@sbg.ac.at

Not only has Brentano's account of consciousness had significant influence in recent years; it also foresaw many of the contemporary debates about the nature of consciousness. Indeed, much of the recent literature on Brentano emerged as part of the work on higher-order theories (HOT) of thought and perception, same-order theories of consciousness, representationalism, intentionalism, and self-representationalism. For all these theories of consciousness and intentionality, Brentano's writings on intentionality and consciousness are often seen to illustrate one aspect or another of the respective theories. Since these theories work with very different assumptions, it might seem that Brentano's conception of consciousness suffers from at least some inconsistencies or, more reasonably, that some of his writings leave room for interpretation. Fisette's paper tries to shed light on Brentano's account of consciousness, and proposes a reconstruction of his view inspired by some of his later ideas on the nature of consciousness and the soul. In what follows, I identify Fisette's real target (sect.1) and challenge his conception of the mental agent as central to Brentano's account (sect. 2 and 3). In section 4, I formulate some doubts about the sources used by Fisette, and in section 5 I propose another reading of the relation between the unity of consciousness and the mental agent in the late Brentano.

The target

Fisette's aim, in this paper, is to criticize a thesis according to which Brentano's views on the mind should be considered along the lines of a higher-order theory of consciousness (T1). Fisette suggests that the 'changes that Brentano brings to his initial theory of consciousness [make it] clear that one may not reduce it to [a] higher-order theory of consciousness'. Furthermore, he points out that, significantly, Brentano never held the view that consciousness was relational (or 'transitive'): 'consciousness represents within Brentano's theory a form of intransitive self-consciousness which is intrinsic to the agent'.

According to Fisette, the interpretation of Brentano's theory of consciousness as a HOT-theory is not only widespread, it is also persistent: it simply 'prevails in Brentanian studies'. This statement is surprising, especially when we consider the authors and papers supposedly championing this interpretation: Güzeldere (1997) doesn't make any statements regarding the specific nature of Brentano's theory of mind (his name is mentioned along with James and Locke, in a list of philosophers who took consciousness to be some kind of perception of a mental state), while Siewert (1998) refuses to commit himself to interpreting Brentano's as a HOT-theory of consciousness. Zahavi (2004) simply underlines structural similarities between Brentano's account and HOT-theories. Textor (2006) does propose an interpretation using some higher-order structures, and both Gennaro (1996) and Janzen (2008) see in

Brentano's account a conception of consciousness as reflective or self-referential, but neither propose interpreting Brentano's theory as a HOT-theory proper. Rather, the common ground that unites these interpretations is simply the view that Brentano's account of consciousness involves a reflective or self-referential moment in every conscious state. This feature is certainly not incompatible with a HOT-friendly theory of consciousness (see for instance, Kriegel 2003), but having this feature doesn't make a theory of consciousness a HOT-theory, and the authors mentioned here can hardly be seen to champion (T1). Who, then, is speaking up for (T1)? According to Fisette, Rosenthal himself would defend (T1): 'Rosenthal (1991, 30, n. 4) nevertheless considers that the heart of the Brentanian theory of consciousness "is virtually indistinguishable from that for which [he] argue[s]"'. Unfortunately, Fisette misquotes his opponent: Rosenthal says quite the contrary: '[Brentano] gives no reason for his insistence that this awareness of conscious mental states is intrinsic to those states; and if it is not [intrinsic], the resulting theory is virtually indistinguishable from that for which I argue below' (ROSENTHAL 1991, p. 30, n. 14). Contrary to Fisette, it seems clear to me that Rosenthal fully realizes that the intrinsicality of consciousness to mental states is a fundamental feature of Brentano's theory of mind. Therefore, attributing (T1) to Rosenthal seems misguided.

What, then, is Fisette's real target? Perhaps the view attacked by Fisette would be better formulated in the following way:

> (T2) Brentano's account of consciousness makes consciousness a relational (or transitive) feature of the mind.

Here, although for different reasons, at least some intentionalists and (self-) representationalists would be sympathetic to (T2).[1] Also, many papers and books published by Brentano himself during his lifetime seem to offer some evidence for (T2).[2] Unfortunately, Fisette neither addresses the intentionalist and self-representationalist readings of Brentano directly, nor comments on Brentano's own texts supporting (T2), but relies on a posthumously published work, edited by Franziska Mayer-Hillebrand in 1954, under the title *Religion and Philosophy* (BRENTANO, 1954) a collection of heavily-edited manuscripts bristling with unmarked personal additions by the editor herself as well as by Alfred Kastil, who undertook preliminary work on this edition in the 1930s. In Fisette's view, the concept of a mental agent (*der psychisch Tätige*) developed in some parts of this book would confirm the non-relational nature of Brentano's account of self-consciousness. Concomitantly, it would 'first attempt to answer the question as to what constitutes the real substrate of the

[1] See for instance Crane (2007) or Kriegel (2013).

[2] See also Fréchette (2011) for further elements in this direction.

complex mental act [...] apprehended in inner perception'. Indeed, Fisette takes the concept of the mental agent (which he also calls 'consciousness de se'), as the bearer of intransitive self-consciousness, to be both Brentano's answer to potential intentionalist or representationalist criticisms and a complement to his theory of mind in the *Psychology*.

The details of this view run as follows: consciousness de se should be seen as a 'new mode of consciousness' thanks to which our intransitive conscious states are said to be conscious. Brentano's theory of consciousness would therefore have three levels: (1) transitive conscious mental states (seeing a blue patch); (2) intransitive conscious mental states (consciously seeing); and (3) consciousness de se (1st-person thought that I am in the process of seeing, which Fisette characterizes as intransitive). Following Fisette's interpretation, levels (1) and (2) were considered by the early Brentano to be parts of the mereological whole that constitutes the unity of consciousness. In Fisette's view, the early Brentano thought that the mereological relation of consciousness, with its parts (1) and (2), was all there was to say about consciousness. But according to Fisette, the late Brentano wasn't satisfied with this model, mainly because (a) 'the nature of the substrate that underlies and unifies as a whole the modes of consciousness' is left untouched by the earlier model; and (b) no details are given in the earlier account 'on the status of the simultaneous consciousness that accompanies the various elements that make up this unity'. Brentano therefore introduced level (3) to address these issues, thereby offering an account of consciousness which is not a full-blown higher-order theory (rather a multi-layer theory), nor a typical same-order theory, nor an intentionalist (or representationalist) model of self-consciousness (or self-representationalist model) although it includes many elements of each of these theories.

The proposal is original and provocative. Unfortunately, Fisette doesn't go into the details of his proposal, which remains speculative to a large extent: from an historical point of view, it falls short of textual evidence supporting the central thesis, according to which consciousness de se (as a substrate) *makes* our intransitive conscious mental states conscious. In fact, as I will suggest, Brentano never doubted that there is a substrate to our conscious mental states. This substrate is called the soul, but *pace* Fisette, in Brentano it never plays any role in the explanation of what makes mental states intransitively conscious. Concerning Fisette's points (a) and (b), I don't see how determining the nature of the substrate would offer an answer to the question of what makes our mental states conscious: the substrate being a brain, a transcendental ego, a person, etc. wouldn't change the fact that simultaneous mental states are co-conscious, i.e. that they belong together as parts of larger whole. The substrate could definitely help answer the question of what makes consciousness

identical over time,[3] but Fisette doesn't explore this possible motivation in Brentano's later account of consciousness.

From the 'psychology without a soul' to the substantial bearer of consciousness

It seems relatively unproblematic to say that Brentano tuck to the thesis that the mental (or 'psychical') is, in some important sense, distinct from the physical. The realm of the mental is immaterial, while the realm of the physical is spatio-temporally extended, i.e. it is material. He also remained firm about the relation between the soul and the mental acts: the soul is a substance, whose accidents are the mental acts. We find this conception in the early Metaphysics lectures from the 1860s, in the *Psychology from an empirical Standpoint* from 1874, and in later manuscripts belonging to the so-called 'reistic' period. Not only did Brentano remain, all his life, true to his faith—he believed in the existence of God and in the immortality of the soul—he also consistently saw the demonstration of these two theses as a crucial part of his philosophical endeavor. This being said, Brentano never brings any assumption about the existence of the soul into play when he discusses the unity of consciousness or any other matter concerning psychology. The main reason for this is that he considers psychology to be a science of experience. Souls are not experienced. Phenomena are:

> If someone says that psychology is the science of the soul, and means by 'soul' the substantial bearer of mental states, then he is expressing his conviction that mental events are to be considered properties of a substance. But what entitles us to assume that there are such substances? It has been said that such substances are not objects of experience; neither sense perception nor inner experience reveal substances to us. (BRENTANO, 1874/1973, p. 8).

In fact, Brentano wished to establish a scientific psychology liberated from metaphysical assumptions about the existence of the soul: 'whether or not there are souls, the fact is that there are mental phenomena'.

Fisette's supposition that this attitude developed into a problem for the late Brentano is not unfounded. Indeed, both Kastil and Kraus make similar observations. In 1924, Kraus goes so far as to put into question Brentano's statement in the *Psychology from an Empirical Standpoint* to the effect that 'there is no such thing as the soul, at least not as far as we are concerned, but psychology can and should exist nonetheless, although, to use Albert Lange's

[3] I discuss the question of the unity of consciousness over time in Brentano in Fréchette (2012).

paradoxical expression, it will be a psychology without a soul' (Brentano 1874/1973, 8). In his 1924 preface to the book, Kraus comments on this phrase:

> That Brentano had no intention of writing a 'psychology without a soul' as is often said should not need to be pointed out. His discussion of the unity of consciousness is an extremely important preliminary to consideration of the problem of the soul. According to Brentano's later theory, words like 'consciousness', 'presentation' and 'judgment' are mere grammatical abstractions which have no independent meaning. However, 'someone with something before his mind' is an independently meaningful expression. In other words, it stands to reason that mental states must have a subject whose accidents they are; furthermore, in conceptualizing ourselves as mental agents, we perceive this subject directly, even if only extremely generally. So the problem of the soul is only a question of *what* is the subject of consciousness and not of *whether* such a thing must exist. (KRAUS 1924, in BRENTANO, 1874/1973, p. 361).

It is true that in 1874 (but also later), Brentano considered discussions on the unity of consciousness as preliminary to reflections on the immortality of the soul. The *Psychology* was originally supposed to include a sixth book that would deal with this topic (BRENTANO, 1874/1973, p. 55). But Brentano never said that the immortality, or even the existence of the soul, was a condition for the unity of consciousness. Following Kraus' view, the late Brentano would have said that the expression 'unity of consciousness' has no independent meaning since 'consciousness' doesn't designate a *realis*. As such, talk of the 'unity of consciousness' should be reduced down to 'unity of someone with something before his mind'. In other words, when one speaks of consciousness, one actually speaks of 'someone with something before his mind'. If this reduction is to be in any way meaningful, the term 'mind' must itself be the designation of a real entity. Following Kraus, this would mean that the unity of consciousness is nothing but the unity of the soul. The consequence of reism is that 'consciousness' designates nothing other than the soul.

Even if we accept this strong ontological consequence for the theory of consciousness, it is still unclear whether the soul, or self-consciousness *qua substrate*, fills a gap in the earlier theory, despite giving an ontological answer to a phenomenological problem. After all, instead of talking about 'consciousness', and preferring 'mental agent' or 'mental activity', the basis of Brentano's account remains, at bottom, unchanged in his later view, as shown by these remarks from 1911:

> In a single mental activity [...] there is always a plurality of references and a plurality of objects.
> As I have already emphasized in my *Psychology from an Empirical Standpoint*, however, for the secondary object of mental activity one does not have to thing of any particular one of these references, as for example the reference to the primary object. It is easy to see that this would lead to an infinite regress, for there would have to be a third reference, which

> would have the secondary reference as object, a fourth, which would have the additional third one as object, and so on. The secondary object is not a reference but a mental activity, or, more strictly speaking, the mentally active subject, in which the secondary reference is included along with the primary one. Although now no infinite regress of mental references *en parergo* can arise, it does not follow that mental activity is to be conceived as something simple. Even when mental references have the same object, they can still be different if the modes of reference are different. (BRENTANO, 1874/1973, p. 215).

For the late Brentano, the mentally active subject includes both the primary reference (my seeing red) and the secondary reference (my being conscious of seeing red). This statement doesn't really differ from the earlier thesis that every conscious act contains a primary and a secondary object. Whether the *bearer* is a self-conscious substance, a brain, or a mental act in its unity doesn't change anything with regard to the mereological relation between the parts. Also, having a substantial bearer of the secondary relation is certainly a change in the theory, but it remains unclear how this substantial bearer is supposed to give us anything substantial about the nature of the unity of consciousness, or at least anything not already provided in Brentano's earlier account.

Why a substantial bearer of consciousness?

Even today, readers and students of Brentano seem unable to identify the deeper motives that led him, around 1904, to reism, namely that one can only present things, i.e. *n*-dimensionally extended substances (through their ontologically dependent accidents), since only such things exist. However, since Brentano believed in the existence and immortality of the soul (a 'zero-dimensional substance'), the reistic assumption can hardly be seen as a change of mind regarding his conception of consciousness and the soul. In other words, even if one accepts Fisette's claim that the introduction of the mental agent changes something in Brentano's general picture of consciousness, we still have to find a reason for this change, since presumably it is supposed to be an improvement on the earlier theory.

I see at least one important reason for this change. Following his reistic turn, Brentano rejected all entities that weren't *realia*. In his earlier view, intentionality was thought to be a relation to an immanent object: an *irrealis*. My imaginings of a unicorn and a horse both have respective intentional objects in the same sense, following this view. Rejecting *irrealia* forced Brentano to review his conception of intentionality as a relation between a subject and an intentional or immanent object. Thus, intentional relations in the earlier view were doomed to be mere *irrealia* after the reistic turn. This seems to me a plausible reason for the late Brentano to reject (T2) and to try to work out a strictly non-transitive account of consciousness. But if this is the

case, the concept of the mental agent as bearer of conscious acts cannot be seen as a complement to the earlier theory; it is a simple consequence of reism. Even if this is the case, it doesn't imply that the mental agent guarantees the unity of consciousness.

In other words, the introduction of the mental agent cannot be interpreted as a sign that Brentano's account of self-consciousness was necessarily intransitive, or that this is expressed in his reism. On the contrary, reism constitutes a break with his earlier account of *irrealia*. From then on, consciousness cannot possibly be explained in intentional terms in a reistic framework. The intransitive substantial self-consciousness advocated in reism is certainly not a natural complement to the earlier theory but is instead part of a very different theory. Brentano himself referred later to his earlier theory of intentionality as his 'old theory', which in his view was superseded by the newer one.

The mess in Brentanian scholarship

These different phases in the evolution of Brentano's thought, together with the doctrinal conflicts that emerged among his students on the appropriate treatment of his posthumous writings, still today constitute a major obstacle to a clear and faithful treatment of Brentano's ideas. The materials used by Fisette for his reconstruction of Brentano's account of consciousness are no exception. The passage from *Religion und Philosophie* is part of an essay entitled 'Über die Geistigkeit und Unsterblichkeit der menschlichen Seele' (*On the Spirituality and Immortality of the Human Soul*). This essay was written by Kastil in 1942, and not by Brentano. Here Kastil tries to give an account of Brentano's 'numerous attempts at giving a proof of the spirituality of the psychical subject' (KASTIL, 1942; BRENTANO, 1954, p. 265). Some of these attempts are inspired by Brentano's lecture on the being of God (*Vom Dasein Gottes*) given in Vienna in 1891/92; other parts of the essay are taken from a lecture by Marty on body and soul. Supposedly even Stumpf's 'Leib und Seele' from 1896 (STUMPF, 1903) was influenced by these lectures.[4] Putting aside the fact that the manuscript in question was not written by Brentano, nothing in the text used by Fisette is actually referable to Brentano's 'late position', since it is composed of and/or inspired by numerous texts by Brentano (and Marty) belonging to different unidentified periods.

[4] Interestingly, Kraus (1924) is stating exactly the contrary when he says that it was Stumpf's lecture of 1896 (Stumpf (1903)) that paved the way for Brentano's alleged change of mind regarding the mental subject (BRENTANO 1874/1973, p. 316).

Unity (and the bearer) of consciousness

Even if one takes the Kastil paper into consideration in Brentano (1954), it is not stated there that the mental agent is what makes the unity of consciousness possible. In fact, the point made here by Kastil is different to that put forward by Fisette. Here, Brentano and/or Kastil are saying that *since* there is something like the unity of consciousness (and with it the unity of both sensory and non sensory phenomena), a so-called 'semi-materialistic' position like Aristotle's—according to which the bearer of the consciousness has to be material to some extent—is not defendable. The nature of the bearer plays no central role in the point made here by Brentano and/or Kastil. Associating, like Aristotle, the sensory experience with a kind of sensitive-material consciousness, is, according to Brentano and/or Kastil, not defensible, since it would allow for different conscious entities—a semi-materialistic position that Brentano and/or Kastil would reject. Even if we set aside the problematic authorship of the text, the position advocated there does not state the necessity of a substantial bearer. Rather, it confirms the earlier account of the unity of consciousness, keeping the same basic assumption that the unity of consciousness—the unity of the mental phenomena—is a primitive fact warranted by inner perception—a primitive fact that is one of the central features in Brentano's distinction between the mental and the physical, and which excludes Aristotle's semi-materialism in favor of a dualist position. Brentano's point in the quote used by Fisette (on semi-materialism) is to 'prove the spiritual nature of the self' (*die Geistigkeit unseres Ich*) and 'definitively refute all materialism' (*dem Materialismus jeder Ausweg entziehen*) (BRENTANO, 1954, 228).

I want to argue that what is introduced in the quote is not a mental agent, but a spiritual self, which is over and above any kind of materialistic conception of subjectivity. The introduction of this 'spiritual self' is not meant to provide a 'deeper' ontological ground, one which would found the unity of consciousness, and nor does it give an account of the status of unifying self-consciousness. In fact, following the text, the unity of consciousness is already a fact secured by inner perception. One might call it the 'mental agent' or the 'basic unifying thing' (*letzteinheitliches Ding*); its ontological nature doesn't play any role in the phenomenological fact of the unity of consciousness:

> [Aristotle] doubly infringes the secured fact of the unity of consciousness. First by conceiving the soul as a composition of corporeal and uncorporeal parts. Second, by attributing to the different parts of our sensory perceptions and desires different parts of the corporeal subject. (BRENTANO, 1954, 224).

> [I]n inner perception, [we are confronted] with *one* basic unifying thing which has a multiplicity of determinations. (BRENTANO 1954, 226).[5]

Later in the same text, the following conclusion is formulated:

> We must think the subject of all our states of consciousness as a non-spatial substance which doesn't constitute a part of the flesh itself, as a spiritual, i.e. zero-dimensional being (*Wesen*). As such, [it is] localized nowhere in the brain, not even in space does it stand locally nearer to a point than to another. For that reason, it can have an immediate effect on every part of the brain and can receive an immediate effect from every part of the brain.[6]

In my view, the account sketched here is quite different from Fisette's reconstruction. Brentano's and/or Kastil's point seems rather to be that the unity of consciousness is what *makes* a being (a creature) conscious. The unity of consciousness is opposed to materialism in this view, since it is the unity of both sensory and non-sensory states that makes a being conscious. In this sense, I would suggest that Brentano shares with Rosenthal the assumption that state consciousness is a primitive fact, and that it explains creature consciousness. Fisette would disagree: following his reconstruction, Brentano should (or wanted to) give an account of state consciousness on the basis of intransitive creature consciousness. I can't see such a project in Brentano's writings. In my view, a Brentanian mental state is conscious because of its *mereological* and *self-referential* structure, and on the basis of this structure alone.

Conclusion

Fisette starts the conclusion of his paper with the following remark: 'Once we consider the changes that Brentano brings to his initial theory of consciousness, it is clear that one may not reduce it to either versions of the higher-order theory of consciousness'.

What I want to show, against Fisette, is that this reconstruction is not attributable to Brentano. A definitive take on Brentano's theory as not being

[5] German original: '[Aristoteles] verstößt gegen die gesicherte Tatsache der Einheit des Bewußtseins, und zwar doppelt, erstens indem er die Seele als Zusammensetzung aus einem körperlichen und einem unkörperlichen Bestandteile faßt, zweitens indem er unsere sinnlichen Wahrnehmungen und Begehrungen Teil um Teil verschidedenen Teilen des körperlichen Subjektes zuweist'; '[I]n der inneren Wahrnehmung [haben wir es] mit *einem* letzteinheitlichen Dinge zu tun, das eine Mannigfaltigkeit von Bestimmungen aufweist'. (p. 226).

[6] German original: 'Wir müssen uns also das Subjekt aller unserer Bewußtseinszustände als eine unräumliche Substanz denken, die nicht einen Teil des Leibes selbst bildet, als ein geistiges, d.h. null-dimentsionales Wesen. Als solches an keiner Stelle des Gehirns lokalisiert, nicht selbst im Raume, steht es keinem Punkte desselben örtlich näher als einem anderen, und kann eben darum auf jeden Teil des Gehirns gleich unmittelbar einwirken und von jedem unmittelbar eine Einwirkung empfangen.' (BRENTANO, 1954, p. 231).

reducible to a higher-order theory of consciousness is certainly not attained here. There definitely are higher-order elements in Brentano's theory of consciousness, as there are elements of a self-representational theory. Even similarities with same-order theories are undeniable. Considering, on top of this, Brentano's complete rejection of materialism, a reconstruction of his theory of consciousness turns out to be a very complicated enterprise.

References

BRENTANO, F. *Religion und Philosophie*. Bern: Francke Verlag, 1954.

_____. *Psychology from an empirical standpoint*. London: Routledge, 1874/1973.

CRANE, T. 'Intentionalism', In: MCLAUGHLIN, B.; BECKERMANN, A.; WALTER, S. (Eds.) *The Oxford Handbook of Philosophy of Mind*. Oxford: Oxford University Press, 2007, p. 474–493.

KRAUS, O. 'Einleitung des Herausgebers', In: BRENTANO, F. (1924). *Psychologie vom empirischen Standpunkt. Erster Band*. Leipzig: Meiner Verlag, 1924, xvii–lxxvi.

STUMPF, C. *Leib und Seele*. Leipzig: Barth, 1903.

FRÉCHETTE, G. 'Brentano über innere Wahrnehmung und Zeitbewußtsein im Zusammenhang mit seiner These zur Einheit des Bewußtseins'. In: DUNSHIRN, A.; NEMETH, E.; UNTERTHURNER, G. (Eds.). *Crossing Borders. Grenzen (über) denken. Beiträge zum 9. Internationalen Kongress der Österreichischen Gesellschaft für Philosophie in Wien*, Vienna, 2012, p. 907-917, 2012. Disponível em: <http://phaidra.univie.ac.at/o:128384>.

_____. « Leibniz and Brentano on Apperception », In: BREGER, H.; HERBST, J.; ERDNER, S. (Dirs.). *Natur und Subjekt. Vorträge 1. Teil , Proceedings of the ninth international Leibniz Congress*, Hanovre, 2011, p. 351-359.

GENNARO, R. *Consciousness and Self-Consciousness*. Philadelphia: John Benjamin Publishing Co, 1996.

GÜZELDERE, G. 'Is Consciousness the Perception of What Passes in One's Own Mind?' In: BLOCK, N. *et al.* (Eds.). *The Nature of Consciousness:* Philosophical Debates. Cambridge: MIT Press, 1997, p. 11–39.

JANZEN, G. *The Reflexive Nature of Consciousness*. Philadelphia: John Benjamins Publishing Co., 2008.

KRIEGEL, U. 'Consciousness, Higher-Order Content, and the Individuation of Vehicles' *Synthese,* v. 134, p. 477–504, 2003.

ROSENTHAL, D. 'The Independence of Consciousness and Sensory Quality', *Philosophical Issues*, v. 1, p. 15–36, 1991.

SIEWERT, C. *The Significance of Consciousness*. Princeton: Princeton University Press, 1998.

TEXTOR, M. 'Brentano (and some Neo-Brentanians) on Inner Consciousness.' *Dialectica*, v. 60, p. 411–431, 2006.

ZAHAVI, D. 'Back to Brentano?' *Journal of Consciousness Studies*, v. 11, p. 66–87, 2004.

Revista de Filosofia
ARGUMENTOS

Joelma Marques de Carvalho*

Franz Brentano's higher-order theories of consciousness

ABSTRACT

This article aims at giving a brief comment on Denis Fisette's interpretation of Higher-Order Theories of Consciousness by Franz Brentano, where consciousness has been seen as a form of intransitive self-consciousness being intrinsic to the agent. In agreement with that interpretation, I want to present a few more basic arguments in order to support that assumption such as, for example, some epistemic thoughts by Brentano given in his books *Psychologie vom empirischen Standpunkte* (1874) and *Die Deskriptive Psychologie* (1982). The present paper has been divided into five sections. The first section deals with the initial understanding of psychology in Brentano. Section two deals with the concepts of consciousness and intentionality. In the third section, the classification of mental phenomena will be presented. Section four refers to the concept of descriptive psychology or phenomenology and finally, I will show the consequences of Brentano's epistemic and ontological arguments related to his concept of consciousness.

Keywords: Philosophy of mind; Brentano; Higher order theory of consciousness; Consciousness; Descriptive psychology.

RESUMO

Este artigo tem por objetivo fazer um breve comentário sobre a interpretação de Denis Fisette das teorias de ordem superior da consciência feitas por Franz Brentano, onde a consciência tem sido vista como uma forma de auto-consciência intransitiva, sendo intrínseca a um agente. De acordo com esta interpretaçao, gostaria de apresentar alguns argumentos básicos para dar suporte àquela assunção, tais como, por exemplo, alguns pensamentos epistêmicos de

* UFC. E-mail: joelma_marques@yahoo.com.br

Brentano dados nas obras *Psychologie vom empirischen Standpunkte* (1874) e *Die Deskriptive Psychologie* (1982). O presente trabalho foi dividido em cinco seções. A primeira seção trata do entendimento inicial a respeito da psicologia de Brentano. A seção dois lida com os conceitos de consciência e intencionalidade. Na terceira secão, será apresentada a classificação do fenômeno mental. A seção quatro se refere ao conceito de psicologia descritiva e à fenomenologia e finalmente, mostrarei as consequências dos argumentos epistêmicos e ontológicos de Brentano relacionados ao conceito de conciência.

Palavras-chave: Filosofia da mente; Brentano; Teoria de ordem superior da consciência; Consciência; Psicologia descritiva.

Psychology

In his work, *Psychologie vom empirischen Standpunkte*, Franz Brentano (1838-1917) presents his theory on consciousness and intentionality. That theory is part of a more general and more ambitious project on the epistemic value of a knowledge which has been generated by pure psychology with respect to other sciences. According to Brentano, psychology doesn´t differ from other sciences due to its methods but due to its research object, that is, its psychological acts. Both mathematics and physiology form the base of psychology, but psychology is thought to primarily rely on internal perception or experience. That is why Brentano entitles his work as psychology from an empiric point of view. The term "empiric" doesn´t refer to those aspects being subject of measurement but to phenomenological or descriptive studies of psychological acts by means of internal experience which is able to produce clear judgments.

"Internal perception" (*innere Wahrnehmung*), however, shouldn´t be understood as an internal observation or insight. Brentano rejects the concept of insight since it is impossible to have an insight or an observation of current psychological acts, because that attempt is prone to modify the mentioned psychological act or even to delete it. Let's take the following example: if someone tries to observe the anger he feels when he listens to the noise of his neighbor's house, his psychological act (to feel anger) could be changed or eliminated at the very moment the person feeling anger is observing that act. Any form of insight as an internal observation of its own psychological acts can only be done in the case of psychological phenomena which aren´t current anymore, such as when we, for example, remember past psychological phenomena. It is only in that sense that we can speak of insights. Yet memory may fail and doesn´t bring about any evidence of internal perception.

An external perception is some kind of perception of bodily phenomena, such as colors, sounds, a landscape we see, and is captured by means of our senses and is observable. Unlike internal perception, external perception doesn´t give us any evidence. That means that, in epistemological terms, judgments of internal perception have to be located on a higher order than judgments of external perception. Inasmuch as natural sciences tend to lean more on external perception than on internal perception, their knowledge shows to be epistemically lower-ordered than the knowledge of psychology.

Consciousness and intentionality

Consciousness has been defined by Brentano as *psychological act*. The term "act" doesn't refer to an activity such as drinking beer or swimming, but to the Aristotelian term "actualita". Thus, he stresses the present and actual features of the psychological phenomena. Still another reason for him to identify "consciousness" and "mental act" is because every psychological act is deliberate and conscious, that means, 1) the content of such an act is an object that is deliberately *inexistent* and refers to an object, and 2) is its own object of internal perception. The expression "inexistence" shouldn´t be understood here as the negation of something´s existence, but as the existence of the referred object "within" the psychological state of the mentioned object. Existence of the intentional or inherent object within the psychological act doesn´t mean the existence in its strict sense since it is merely a deliberate existence (as represented object). Besides, there doesn´t exist any physical object to show that feature. That is the reason why the basic feature of consciousness or of psychological phenomena is the intentional inexistence. Therefore he is making the case of two intentional arguments:

(1) Every psychological act is intentional
(2) Only psychological acts are intentional

The combination of those two arguments became known as "Brentano´s argument". Within that context he distinguishes between two types of consciousness: (i) *primary consciousness* and (ii) *secondary consciousness*. Let´s think about the case that a particular individual A is listening to a particular sound ***x***. In such a case the psychological act of hearing *a* directly refers to the sound ***x*** and to the psychological act, *a* holds the sound ***x*** by means of an intentional inexistence (the sound "exists" within *a*). Primary consciousness comprises the relation between the psychological act of listening, *a*, and its intentional object ***x***, which could be a transcendent object as well as something imaginary. That means that the intentional "relation" is unable to ensure real existence of the referred object. (In fact, it is not a relation

stricto sensu. When "a R b" is meant to be true for "R", any real relation, the individual constants "a" and "b" should designate existent things; nevertheless, such as Brentano accurately noted in the case of a "quasi-relation" of intentionality, the imagination of a centaur only presupposes the existence of a bearer of the psychological act).

When listening to the sound *x*, the individual A doesn´t only have primary consciousness, but also a secondary consciousness of the psychological act *a*, once the psychological act of listening can be an object of its internal perception too. Secondary consciousness encompasses the relation between the psychological act of listening *a* and itself, bearing in mind the internal perception of that individual. The psychological act of listening *a* is deliberately directed to itself. Thus, intentionality and consciousness are *inextricably linked*. That kind of consciousness cannot be understood, though, as consideration or introspection as there is no other psychological act *b* that may refer to the mentioned psychological act *a*. Therefore, Brentano, following inspiration in Aristotle, tries to avoid a return to infinity, because if the existence of a further psychological act *b*, being addressed to the psychological act *a*, should be necessary in order to turn the psychological act *a* into a conscious state, then the existence of a psychological act *c*, heading to the psychological act *b*, should be necessary and so on. Brentano, in his mereological analysis of consciousness, assumes primary consciousness and secondary consciousness as "parts" or "divisions" of consciousness, being without number distinction but just being mentally different. The following remarks are to foster a better understanding of that kind of consciousness.

The classification of mental phenomena

According to Brentano the psychological acts can be sorted on: (i) representations (Vorstellungen), (ii) judgments (Urteile) and (iii) emotions (Gemütsbewegungen). That hierarchical separation is crucial because the psychological acts are cumulative. That means that if an individual A feels an emotion regarding x, then he also has a judgment of x as well as a representation of x. If an individual A has a judgment about x, he also has a representation of x, but it could also be the case that he only has a representation of x but neither judgment nor emotion addressed to that object.

The basic acts of consciousness are the *representations*, since they exhibit and bring the intentional object to the level of consciousness. They are epistemically neuter inasmuch they don´t imply any opinion or any kind of agent judgment about the intentional object. The individual has a representation whenever something arises to the level of consciousness, whether by means of his senses or by means of his imagination.

Judgments involve the individual´s judgment, in other words, at least the acceptance or rejection of existence of the intentional object. As judgments do not add any intentional object to consciousness, representations and judgments are just different ways of consciousness of the same intentional object. Judgments can be classified into a) *apodictic* and b) *assertoric* ones. Apodictic judgments are true judgments within all possible worlds, for instance such as mathematical statements. Assertoric judgments can be true or false, such as for example all judgments of external perception and some judgments of internal perception. The latter can be classified into: (i) *blind* judgments and (ii) *evident* judgments. Blind judgments are those from external perception. For example: when an individual happens to have a visual experience from an object x, then he experiences the representation of x as well as the blind judgment about it, since external perception alone is unable to ensure the existence of the intentional object. Evident judgments are either judgments of internal perception or of secondary consciousness, because in that case a distinction between psychological act and intentional object cannot be done. That means that there is a *real identity* between the represented object and the act representing the object.

Emotions do not include only feelings like love and hatred, but also wishes, intentions, fears etc. In most cases, the intentional object happens to be at the same time object of representation and the judgment presupposed by an emotion, although emotion is sometimes just referring to the corresponding psychological act. Let´s take the following example: When an individual A listens to a sound x, he is going to have the representation of x and a judgment on x, yet the related positive or negative emotion possibly refers only to the psychological act of listening and not just to sound x. In that case the emotion would be just a kind of secondary consciousness.

The coupling of those three kinds of consciousness (representation, judgment and emotion) sometimes has been called *internal perception* by Brentano. Furthermore, those three types of psychological acts aren´t really different from each other from the point of view of mereology, but only in mental terms. That means that when an individual A experiences a visual perception of an object x, that individual doesn´t have in number three psychological acts, but only one single psychological phenomenon, presenting, however, three distinct aspects, namely a representation, a judgment and an emotion.

Descriptive psychology

In his work *Descriptive Psychology*, Brentano splits psychology into *psychognosia* and *genetic psychology*. Psychognosia is the same as descriptive psychology or descriptive phenomenology. It is dealing with "parts" or elements of consciousness, in other words, with psychological acts themselves. Genetic

psychology considers the origins and conditions of psychological phenomena. Those two areas are thought to complete each other. Nevertheless, it is ultimately genetic psychology that rather presumes the study of descriptive psychology than vice versa. On the field of psychognosia it´s impossible to distinguish between appearance and reality, because psychological acts usually seem to us the way they really are. For that reason, Brentano keeps on supporting the thesis that knowledge of descriptive psychology comprises a higher epistemic value then knowledge of natural sciences.

It is in that context that he differentiates between *implicit consciousness* (*awareness in a wider sense)* and *explicit consciousness (awareness in a narrow sense)*. Implicit consciousness occurs whenever the individual is apprehending an object yet doesn´t understand it. Explicit consciousness, on the other hand, occurs when the individual not only has the apprehension or experience of an object but also the act of noticing it (*Bemerken*). For example: while looking at the sky, an individual A is seeing a black spot, although he actually doesn´t notice that spot. In that case, when an individual B asks him what it is the other individual will be unable to answer. So, we may claim that individual A only has implicit consciousness of the object. Moreover, we cannot affirm that there has been any error there since when someone doesn´t perceive something that doesn´t mean he has done something wrong there. An error can only arise when the objects of external perception have been fixed and generalized in a wrong manner. In case of an implicit consciousness, the individual has an experience of the object together with an appreciative judgment (anerkennendes Urteil).

Final Remarks

The final remarks on his seminar in Vienna about descriptive psychology bring about some clarification on his theory of consciousness, because when an individual A only experiences implicit consciousness of an intentional object x he doesn´t have any knowledge about it. Thus, awareness of an object x, be it either a bodily or a psychological phenomenon, doesn´t imply neither knowledge nor a form of reflection about the respective intentional object. The verdict p: "I am aware of x" cannot be replaced by q: "I know that I am experiencing x." In the case of explicit consciousness, though, a verdict p can be replaced by the verdict q. Both kinds of consciousness are feasible occurrences within an internal perception as well as within an external perception.

On one side it is only internal perception or secondary consciousness that is able to produce evident judgments, because there won´t be any doubling of the psychological act, that is, psychological act and secondary consciousness are identical. If there were any doubling of the intentional object in secondary consciousness, that would require some form of internal

observation or insight, something that Brentano declines. Furthermore, he would be unable to assert epistemic argumentation on the fact that descriptive psychology will bring about reliable knowledge. That means that a judgment produced by explicit consciousness of a psychological act will always be evident. In the case of an implicit consciousness of a psychological act, on the other hand, there won´t be a judgment about that psychological act. That means that we might have a psychological act without experiencing knowledge about it. As it won´t be the same thing whether we don´t perceive or whether we make a mistake, that fact won´t jeopardize the status of epistemic value of the internal perception.

To make it short, I agree with Fisette´s interpretation that the epistemic theses upheld by Brentano in his theory of consciousness are shown to be only partly Cartesian and that they come close to Aristotle´s position, whereas his ontological theories are actually Aristotelian. These remarks thus confirm Fisette´s view that Brentano´s consciousness is a form of intransitive self-consciousness which is intrinsic to the agent, or is a pre-reflective self-awareness in an intransitive sense. However, unlike Fisette and according to Brandl, I believe that the pre-reflective theory of self-consciousness has been already present in *Psychology.*

References

BRANDL, Johannes. *Innere Wahrnehmbarkeit und intentionale Inexistenz als Kennzeichen psychischer Phänomene*. Brentano Studien, IV. Verlag J. H. Roll.P. p. 131-153, 1992/1993.

BRENTANO, Franz *Psychologie vom empirischen Standpunkte* 1874. *Von der Klassifikation der psychischen Phänomene* (1982). Band 1. Ontos Verlag, 2008.

_____. *Die Deskriptive Psychologie*. Introduction by Roderick M. Chisholm and Wilhelm Baumgartner. Felix Meiner Verlag, 1982.

_____. *Wahrheit und Evidenz. Erkenntnistheoretische Abhandlungen und Briefe*. Introduction by Oskar Kraus. Felix Meiner Verlag, 1974.

CARVALHO, M. Joelma. *Intentionalitätstheorie beim frühen Brentano und bei Searle*. Philosophia Verlag, 2013.

CRANE, Tim. *Brentano's concept of intentional inexistence*. University College London. Disponível em: <http://web.mac.com/cranetim/Tims_website/Online_ papers_files/Crane%20on%20Brentano.pdf> Acesso em: 2006.

CRANE, Tim. *Intentionalität als Merkmal des Geistigen*. Sechs Essays zur Philosophie des Geistes. Translation by Simone Ungerer and Markus Wild. Fischer Taschenbuch Verlag, 2007.

CHRUDZIMSKI, Arkadiusz. *Die ontologie Franz Brentanos*. Kluwer Academic Publishers, 2004.

FISETTE, Denis. *Franz Brentano and Higher-Order Theories of Consciousness*. *Revista Argumentos*, in this issue.

JACQUETTE, Dale. *Brentano's Concept of Intentionality*. The Cambridge Companion to Brentano, 2004, p. 98-130.

KENNT, Otis T. *Brentano and the relational view of consciousness*. _____. *Man and World*, v. 17. p. 19-51, 1984.

MAREC, J. Christian. Psychognosie und Geognosie. *Apriorisches und Empirisches in der deskriptiven Psychologie Brentanos*. Brentano Studien. Internationales Jahrbuch der Franz Brentano Forschung. Band II. P. 53-61, 1989.

MCALISTER, L. Linda. *Brentano's Epistemology*. In: JACQUETTE, Dale (Ed.). *The Cambridge Companion to Brentano*. Cambridge University Press, 2004, p. 149-167.

MÜNCH, Dieter. *Intention und Zeichen. Untersuchungen zu Franz Brentano und zu Edmund Husserls Frühwerk*. Suhrkamp, 2004, p. 35-78.

VOLPI, Franco *War Franz Brentano ein Aristoteliker? Zu Brentanos und Aristoteles' Konzeption der Psychologie als Wissenschaft*. Brentano Studien. Internationales Jahrbuch der Franz Brentano Forschung. Band II. p. 13-30, 1989.

ARGUMENTOS
Revista de Filosofia

Maria Eunice Quilici Gonzalez*, Mariana C. Broens**

On Denis Fisette's "Franz Brentano and higher-order theories of consciousness": a view from the complex system perspective

ABSTRACT

The aim of the present commentary on Denis Fisette's article *"Franz Brentano and Higher-Order Theories of Consciousness"* is to discuss his account of Brentano's principle of the unity of consciousness from the Complex Systems perspective. Initially a summary of Fisette's writings on Brentano's principle of the unity of consciousness is presented. Hypotheses of the Complex Systems Theory are, then, presented in order to provide foundations for an informational interpretation of Fisette's *complexity problem*.

Keywords: Philosophy of mind; Brentano; Consciousness; Complex systems theory.

RESUMO

O objetivo do presente comentário sobre o artigo de Denis Fisette "Franz Brentano and Higher-Order Theories of Consciouness" é discutir sua explicação do princípio da unidade da consciência de Brentano sob a perspectiva de sistemas complexos. Inicialmente, é apresentado um sumário dos escritos de Fisette sobre o princípio da unidade da consciência de Brentano. Hipóteses da Teoria de Sistemas Complexos, são, então, apresentadas para fundamentar uma interpretação informacional do problema da complexidade de Fisette.

Palavras-chave: Filosofia da mente; Brentano; Consciência; Teoria de sistemas complexos.

* University of São Paulo State. E-mail: gonzalez@marilia.unesp.br
** University of São Paulo State. E-mail: mbroens@marilia.unesp.br

Introduction

In the present commentary on Denis Fisette's article "*Franz Brentano and Higher-Order Theories of Consciousness*", we are going to focus on his fresh account of Brentano's principle of the unity of consciousness.

As Fisette stresses, several difficulties arise from Brentano's view on the unity of consciousness; one that is of particular interest here concerns the difficulty of explaining the nature of the objects of conscious experience. In this context, we are going to discuss Fisette's lucid interpretation of Brentano's principle of unity of consciousness from a provisory *informational* perspective grounded upon hypotheses of Complex Systems Theory (GERSHENSON *et al*., 2007; MORIN, 1982; MITCHEL *et al*., 2002; JUARRERO, 2002; HAKEN, 1983, 2000; BAK, 1996). Special emphasis will be given to the *dispositional* nature of informational relations created between physical and non-physical objects. Due to their own peculiar nature, informational relations are not material, but they may entangle a myriad of nested physical elements belonging to the domain of complex (probably self-organized) systems. We are going to provide reasons to support the hypothesis that given the dispositional nature of informational relations, they may constitute a common element that under certain conditions can unify the objects of conscious experience in complex biological systems. It is hoped that this hypothesis could complement Fisette's interpretation of Brentano's perspective on the unity of consciousness.

The aim here is to discuss the nature of the objects of conscious experience from the Complex Systems perspective. The text is organized into three sections, the first of which summarises our understanding of Fisette's writings on Brentano's principle of the unity of consciousness. In the second section, we introduce the main premises of Complex Systems Theory, providing foundations for our informational interpretation, proposed in the third section, of what Fisette calls *the complexity problem*.

Denis Fisette's account of Brentano's principle of the unity of consciousness

One of the central topics analysed by Fisette is Brentano's principle of the unity of consciousness. According to this principle, conscious experience is not constituted by an aggregate of isolated parts, but comprises a whole integrated unity. As Fisette points out, the parts or *divisives* that constitute the conscious experience "[...] stand in a relation of dependence to the whole" (p. 24). He illustrates Brentano's thesis according to which "every mental act is conscious and includes the consciousness of itself." (p. 15). As an example, he considers the act of hearing a sound and the consciousness of hearing the sound, which are parts of the subject's same, integrated, conscious experience. In the above

example, Fisette stresses that according to Brentano, mental acts have a "double object", namely *primary* and *secondary objects*, which constitute the unified experience of hearing a sound. Here, the primary object is the sound, and the secondary object is the mental phenomenon, which characterizes the experience of hearing a sound.

Brentano's principle of the unity of consciousness, as mentioned, raises several questions, one of which is the difficulty of explaining the relationship between the conscious experience itself and the consciousness of having this conscious experience. Fisette presents three traditional approaches to this question in contemporary Philosophy of Mind (p. 21-23). The first, proposed, for example, by Kriegel (2003), suggests that the primary object of consciousness is *represented* by the secondary object. This approach is formulated by Fisette as follows:

> For any mental state M of a subject S, there is necessarily a mental state M^* such that S is in a state M^*, where M^* represents M, and $M^* = M$. (p. 22).

The second approach, proposed by several advocates of higher-order theories of consciousness (represented, in Fisette's analysis, by Rosenthal), presupposes that there is a numerical distinction between lower and higher level conscious states. Fisette summarizes this approach as follows:

> For any mental state M of a subject S, there is a mental state M^* such that S is in the state M^*, where M and $M^* \neq M$. (p. 22).

Finally, the third approach focuses on a *mereological* relation between the primary and secondary objects, both considered as parts of a whole. This whole/part kind of connection is expressed by Fisette (p. 23) as:

> M^* = Representation of the primary object
> M^{**} = Representation of the secondary object
> M = The whole (or complex) unifying M^* and M^{**}
> For any mental state M of a subject S, M is conscious iff there is a M^* and a M^{**}, such that (i) M^* is a part of M, (ii) M^{**} is a part of M, and (iii) M is a whole which M^* and M^{**} are parts of.

Fisette considers that this third view, on the relationship between the conscious experience itself and the consciousness of having this conscious experience, contemporarily developed by van Gulick (2006), amongst others, is shared by Brentano, especially in his later writings. In this sense, he argues that "[...] the consciousness of the primary object and the consciousness of the secondary object are metaphysical parts or, in Brentano's words, divisives that belong to one and the same phenomenon." (FISETTE, 2015, p. 23).

In short, a fundamental aspect of Brentano's theory of consciousness, coherently analysed by Fisette, is the thesis that primary and secondary objects of consciousness are interdependent and constitute a unity. This thesis gives place to what Fisette calls *the complexity problem,* which is: "… the problem of unifying within inner consciousness the entire complex of elements involved in the constitution of our mental life" (p. 24). In what follows, we are going to investigate this problem from the informational perspective in the context of Complex Systems Theory.

A complex systems approach to the nature of the objects of consciousness: any contribution to Fisette's complexity problem?

In his inspiring 1948 paper "Science and Complexity", Warren Weaver proposes a classification of scientific problems into three main categories:

1. Problems of simplicity: Those problems that can be described and solved in terms of two or a few fixed variables.

2. Problems of disorganized complexity: Problems involving numerous variables, whose solutions (if they exist) require probability analysis.

3. Problems of organized complexity: Those problems involving a moderate number of variables and dynamic relations that cannot be solved only by means of probability analysis.

Weaver (1948) stresses that Type 1 problems were successfully investigated and solved during the seventeenth, eighteenth, and nineteenth centuries, guiding great progress in the domain of the physical sciences. Investigations of this type of problem led to the invention, for example, of the telephone, automobiles, and diesel engines, amongst others, but there were clear limitations in the study of biological, psychological, medical, and social problems.

In the nineteenth and early twentieth centuries, Type 2 problems (of disorganized complexity) were investigated in the areas of thermodynamics, logic, mathematics, and aspects of economics involving numerous variables, by means of probability analysis.

It was only in the twentieth century that Type 3 problems (of organized complexity) were investigated. These problems involve a moderate number of variables, and their main characteristic is the dynamic dependency relations that are established in the *communication* amongst members of a self-organized system. The self-organized character of these dynamic interrelations cannot be satisfactorily described only in terms of the probability statistics that seems to be adequate for the analysis of Type 2 problems.

As suggested by Weaver, the power of computers in dealing with information processes, and the interdisciplinary collaboration amongst

researchers in different areas, opens up a promising new perspective for understanding Type 3 problems of organized complexity. In this context, the novelty of the present exploratory commentary is the indication of a possible way of conceiving Fisette's view of Brentano's principle of the unity of consciousness (considered here as a Type 3 problem) from an informational perspective, grounded on Complex Systems Theory hypotheses. In general, the analysis of this theory involves the use of a number of mathematical formulae, which will be left aside in the present case, given that our main interest is to discuss the conceptual presuppositions of the theory. A complex system can be defined as:

> […] [an] organization which is made up of many interacting parts […] In such systems the individual parts - called 'components' or 'agents' - and the interactions between them often lead to large-scale behaviours which are not easily predicted from a knowledge only of the behaviour of the individual agents. (MITCHEL & NEWMAN, 2002, p. 2).

From the perspective of the Theory of Complex Systems, interactions amongst elements at the microscopic level may produce the emergence of order parameters at the macroscopic level of a self-organizing system. Order parameters can be understood here as emergent informational patterns that express several levels of dependency amongst elements on different scales. As Haken (2000) argues, when order parameters emerge, they subjugate the behavior of the individual elements that have generated them, producing new characteristics at the macroscopic scale (the term "order parameter" is used here, in a technical sense, to indicate the emergent structuring property of a complex informational system). In the case of living systems, under certain conditions, changes at the microscopic level may initiate the emergence of informational patterns that could, in turn, create new informational patterns at the macroscopic scale.

The following two basic properties of complex systems are of special interest here: (a) self-organization, and (b) the holographic principle. Self-organization can be characterized as a process through which new forms of organization emerge solely from the dynamic interaction amongst elements - initially independent - without any *a priori* plan or central controller. This process can be developed in primary or secondary ways (ASHBY, 1962; DEBRUN, 2009), described by Gonzalez & Haselager (2005, p. 7) as follows:

i) *Primary self-organization* involves the encounter between organic or inorganic elements, initially separated (or with independent behaviors). These elements get together […] initiating a spontaneous interaction amongst themselves in such a way as to give place to structures or distinct forms of organization, without a central controller;

ii) *Secondary self-organization,* in turn, happens when under certain circumstances there appear disturbances that provide sufficient conditions for the system that is primarily self-organized to learn how to adjust the

communication amongst its element, creating new stable patterns or order parameters that may control the system.

We understand that both primary and secondary self-organization can constitute the core of organized complexity. If it happens that the holographic principle applies to living self-organized systems, then they may be able to express the unified interactions between their constituent elements at the micro- and macroscopic scales. Morin (2001, p. 150) describes the holographic principle according to which " [...] not only its part is in the whole, but the whole is also in each part."

To conclude the present paper, we indicate the role played by informational patterns in the whole/parts dynamic that is implicit in the holographic principle.

An informational approach to the principle of the unity of consciousness

There is no consensus about the proper characterization of the concept of information in contemporary studies, but most researchers emphasise the *relational* nature of information that comprises the interdependence between actions, events, and messages, amongst others. Inspired by Shannon & Weaver's Mathematical Theory of Communication (1949), Dretske (1981) characterizes information as an indicator of relations that exists objectively in the world. In this sense, given two interdependent events, the occurrence of one provides an amount of information about the occurrence of the other. In contrast, an aggregate of independent events provides no information about their occurrences.

Thus, informational relations necessarily express a conditional property, but they differ from causal relations in that the first involves chance and the possibility of choices. Dretske (1981, p. 20) describes the distinction between causal and informational relations as shown in Figures 1 and 2. Figure 1 illustrates a direct one-way link between the occurrence of the state $s2$ in a source and the state $r2$ in a receptor. In contrast, Figure 2 indicates the many possibilities that resulted in the connection between $s2$ and $r2$.

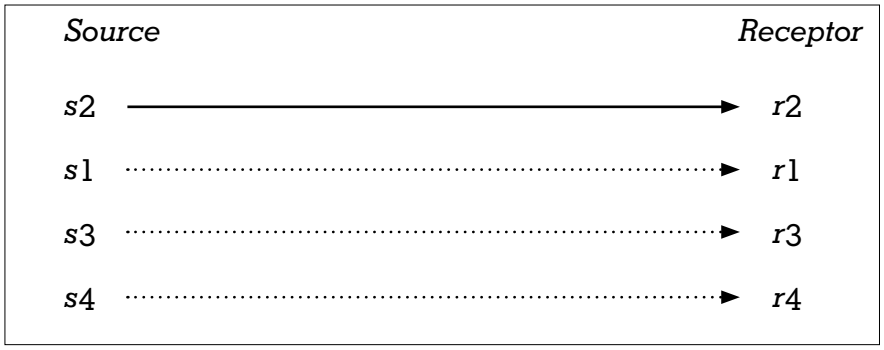

Figure 1 - Diagram of a causal relation, as depicted by Dretske (1981, p. 28).

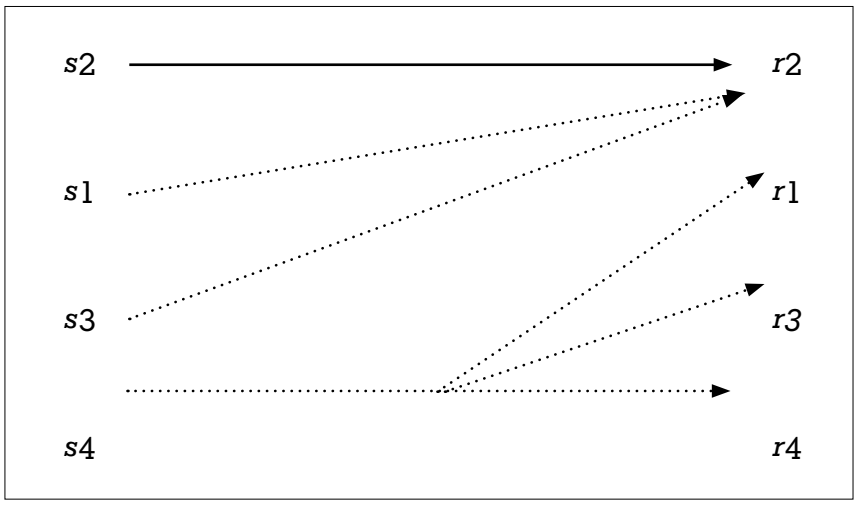

Figure 2 - Diagram of an informational relation, as depicted by Dretske (1981, p. 28).

The difference between informational and causal relations allows *meaning* to be developed in the first type of relation, but not in the second. Considering this important distinction between causal and informational relations, our hypotheses are that:

(i*) Information, characterized as ecological "invariant" features of the world, may constitute the basic elements of the primary object of consciousness;

(2i*) Invariant features of the world may give place to affordances, in that they have the potential to enable organisms to encounter opportunities for action (GIBSON, 1986; TURVEY, 1992).

(3i*) In complex biological systems, meaningful information emerges in consciousness as a result of the agent's adaptive interaction with the environment.

Even though hypothesis (2i*) belongs to Ecological Psychology, and is well known for its anti-representational view of perception-action, we understand that it can be "hired" here to describe the basic informational interaction between agent and environment.

In a related way, Dretske (1992) and Adams (2003) propose the concept of natural *meaning,* understood as an indicator of events in the world, to explain the basic informational relation established between agent and environment: in a certain environment, smoke naturally means, or indicates, fire. In contrast, non-natural meaning is what they call *genuine meaning*, which involves systemic reasons and learning. According to Adams (2003, p. 475-476): "… the word smoke does not naturally mean or indicate fire, but it does semantically mean *smoke*". We understand that information with *genuine meaning* could be a candidate for illustration of Brentano's secondary object of consciousness.

Considering the suggestion of Dretske and Adams that meaningful information can be described in terms of natural and genuine senses, our provisional hypothesis is that both primary and secondary objects of consciousness can be understood as having the same informational nature, despite their specific differences. If this hypothesis is acceptable, then Fisette's *complexity problem,* concerning the unification of the elements involved in the constitution of our mental life, could be investigated from the informational perspective enriched by Complex Systems Theory.

Final comments

To conclude this provisional commentary, we are going to indicate possible contributions of Complex Systems Theory to the analysis of Fisette's *complexity problem,* considered from an informational perspective, as outlined in Section III.

In his inspiring systemic approach to information, Bateson (2000) argues that information is *the difference which makes a difference* to organisms. In his perspective, differences do not make any difference to stones, artefacts, and even machines. According to him, the biological world encompasses nested relations - patterns of information - that provide dynamic organizations, some of which are shared amongst all living beings. However, not all patterns of information constitute objects of consciousness, and this seems to require the perception of relevant differences in the context of action.

From our perspective, in complex biological systems information exchanged among the communicating parts allows self-organizing processes to be established on different scales. By means of secondary self-organization (as indicated in Section II), dispositions may be created and developed in the form of habits and abilities that allow the establishment of constraints for thought and action.

Given the *dispositional* nature of informational relations that may be created between physical and non-physical objects of complex biological systems, we suggest that they constitute a common element that under certain conditions can unify the objects of conscious experience. This hypothesis could complement Fisette's interpretation of Brentano's view on the unity of consciousness. Furthermore, the holographic principle, indicated in Section II, could help in addressing Fisette's *complexity problem*: "[...] the problem of unifying within inner consciousness the entire complex of elements involved in the constitution of our mental life." (p. 24).

In summary, we have here considered information as a self-organizing process of pattern formation that allows the establishment of conditional dispositions in complex biological systems. In these systems, high-level informational structures might emerge that have the ability to create and

change habits through secondary self-organization. These high-level informational structures may well be seen as the secondary objects of consciousness, emergent from the interaction amongst primary objects (which are also informational patterns, with different structures). In dynamic communication, both of them may produce unified conscious experience. From this perspective, a conscious experience is not constituted by an aggregate of isolated parts, but comprises an integrated whole of informational patterns that communicate on different scales.

References

ASHBY, W. R. Principles of the self-organizing system. In: *E:CO special double issue.* v. 6, n. 1-2, 2004, p. 102-126, 1962, (Classical Papers).

BAK, P. *How nature works:* the science of self-organized criticality. New York: Copernicus, 1996.

BATESON, G. *Mind and nature:* a necessary unity. New York: Cambridge University Press, 1979.

_____. *Steps to an ecology of mind*. Chicago: Chicago University Press, 2000.

DEBRUN, M. *Brazilian national identity and self-organization*. Campinas-Brazil: UNICAMP University Press, CLE Collection, 2009.

DRETSKE, F.I. *Knowledge and the flow of information*. Oxford: Basil Blackwell Publisher, 1981.

_____. *Explaining behavior: reasons in a world of causes*. Cambridge, Mass.: MIT Press, 1992.

DRETSKE, F.I. *Naturalizing the mind*. Cambridge, Mass: MIT Press, 1995.

GERSHENSON, C., AERTS, D., EDMONDS, B. (Eds.). *Worldviews, science, and us:* philosophy and complexity. Singapore: World Scientific, 2007.

GIBSON, J.J. *The ecological approach to visual perception*. Boston: Houghton Mifflin, 1986.

GULICK, R. Van. Mirror, mirror - is that all? In: KRIEGEL, U.; WILLIFORD, K. (Eds.). *Self-representational approaches to consciousness*. Cambridge: MIT Press, 2006, p. 11-39.

HAKEN, H. *Synergetics*. Berlin: Springer-Verlag, 1983.

_____. *Information and self-organization*. New York: Springer-Verlag, 2000.

HASELAGER, W.F.G., GONZALEZ, M.E.Q. Creativity: surprise and abductive reasoning. *Semiotica*, v.1, n. 4, p. 325-341, 2005.

JUARRERO, A. *Dynamics in action:* intentional behavior as a complex system. Cambridge, Mass.: MIT Press, 2002.

MITCHELL, M., NEWMAN, M. Complex Systems Theory and evolution. In: PAGEL, M. (Ed.). *Encyclopedia of Evolution*. New York: Oxford University Press, 2002. Available at: http://web.cecs.pdx.edu/~mm/EncycOfEvolution.pdf. Access: 23 may 2013.

MORIN, E. Notas para um Emilio contemporâneo. In: PENA-VEJA, A., ALMEIDA, C.R.S., PETRAGLIA, I. *Edgar Morin:* ética, cultura e educação. São Paulo: Cortez, 2001.

SHANNON, C., WEAVER, W. *A mathematical theory of communication*. Urbana: University of Illinois Press, 1998. (first edition: 1949).

TURVEY, M.T. Affordances and prospective control: an outline of the ontology. *Ecological Psychology*, v. 4, p. 173–187, 1993.

WEAVER, W. (1948). Science and complexity. *Classical papers – science and complexity*, v. 6, n. 3, 2004.

ARGUMENTOS
Revista de Filosofia

Paul Bernier*

Brentano's 'revised' theory consciousness

ABSTRACT

Three substantial issues raised by Fisette's interpretation of Brentano's views on consciousness are discussed. The first concerns the difference between "transitive" and "intransitive" consciousness. The second concerns what Fisette proposes as Brentano's *revised* theory of consciousness, where the notion of a mental agent as a "unified real being" plays a central role. This notion is rejected and some alternative interpretations, which are in the spirit of Brentano's theory, are proposed and defended. Finally, it is pointed out that Fisette's interpretation remains unclear as to whether Brentano's view is compatible or not with Rosenthal's transitivity principle. I argue that while Brentano's revised theory is not intentionalist, as Fisette makes it clear, it is nonetheless compatible with the transitivity principle, contrary to what Fisette claims.

Keywords: Philosophy of mind; Brentano; Transitivity principle; Intentionalism; Consciousness.

RESUMO

Três problemas substanciais levantados pela interpretação de Fisette a respeito das visões de consciência de Brentano são discutidos. O primeiro é concernente à diferença entre consciência "transitiva" e "intransitiva". O segunda trata do que Fisette propõe como sendo a teoria revisada da consciência de Brentano, onde a noção de agente mental como um "ser real unificado" desempenha um papel central. Esta noção é rejeitada e algumas interpretações alternativas, que estão no espírito da teoria de Brentano, são propostas e defendidas. Finalmente, é apontado que a interpretação de Fisette permanece pouco clara a respeito da questão se a visão de Brentano é compatível ou não com o princípio da transitividade de Rosenthal. Argumento que enquanto a teoria revisada de Brentano não é intencionalista, como Fisette deixa claro, é, entretanto, compatível com o princípio da transitividade, contrário ao que Fisette reinvidica.

Palavras-chave: Filosofia da mente; Brentano; Princípio de transitividade; Intencionalismo; Consciência.

* Université de Moncton. E-mail: paul.bernier@umoncton.ca

In the target paper, Fisette presents a detailed interpretation of Brentano's views on consciousness as they are found in *Psychology from an Empirical Standpoint* and as they have evolved in some of his posthumously published works. Fisette's endeavour is well motivated in light of the influence of Brentano on important contemporary theories of consciousness which fall, broadly speaking, under what is known as higher-order approaches to consciousness. The paper is dense and it raises many interesting issues, either exegetical issues or substantive philosophical issues with respect to contemporary debates on the nature of consciousness. In what follows, I focus on three issues of the latter kind. The first concerns the difference between "transitive" and "intransitive" consciousness as these notions are used by Fisette throughout the paper, particularly in his interpretation of Brentano's two theses on consciousness (see section 4). The second concerns what Fisette proposes as Brentano's revised theory of consciousness which he presents towards the end of the paper (sections 7 and 8), where the notion of a mental agent, understood as a "unified real being", plays a central role. It is not my purpose to criticize Fisette's interpretation on exegetical grounds. I take his interpretative hypothesis at face value, but I try to clarify what Brentano's revised theory of consciousness amounts to, especially with respect to the question of the relationship between the primary and secondary object of mental states. I point out that Brentano's revised theory is interesting because it suggests a way to make sense of the strong intuition that conscious mental states necessarily involve a sense of "for-me-ness", as it has recently been stressed in the literature (LEVINE 2006, and KRIEGEL 2009). Brentano's notion of a mental agent as a "unified real being", however, strikes me as something implausible. So I suggest some alternative interpretations which are in the spirit of Brentano's theory but which come short of postulating a mental agent, in Brentano's sense. I argue that these alternative interpretations are more plausible than Brentano's revised theory. Finally, in the last section, I point out that Fisette's interpretation remains unclear as to whether Brentano's view is compatible or not with the transitivity principle, which is central to Rosenthal's higher-order thought (HOT) theory. I argue that while Brentano's revised theory is not intentionalist, as Fisette makes it clear, it is nonetheless compatible with the transitivity principle.

"Transitive/Intransitive"

Fisette's discussion of the similarities and important dissimilarities between Brentano's theory of consciousness and Rosenthal's HOT theory rests crucially on his use of the distinction between so-called transitive and intransitive consciousness. At several places, he claims that according to

Brentano's theory, consciousness turns out to be intransitive, for instance when at the end of the paper he states that "[…] consciousness represents within Brentano's theory a form of intransitive self-consciousness which is intrinsic to the agent." (p. 32). In section 4, where he discusses Brentano's two theses, Fisette argues in support of his interpretation of Thesis II, namely "2b. Every mental phenomenon is an object of consciousness." (p. 17), on the grounds that the alternative interpretation according to which "is conscious" is used in an intransitive sense would "stand in contradiction with Thesis I", according to which "1b. Every mental phenomenon is consciousness of something". Fisette's use of the notions of transitive and intransitive consciousness, however, is puzzling especially because this is not a distinction found in Brentano, but one which was only recently introduced in the literature (ROSENTHAL, 1986, 1997 and TUGENDHAT 1979).

In some important passages, Fisette uses this distinction in the material mode as if "transitive" and "intransitive" would denote some properties of consciousness, even suggesting that they are incompatible properties of mental states. Some of his uses of this distinction also suggest that it would capture a distinction between rival views about consciousness, for instance a view according to which consciousness is essentially transitive and a view according to which it is essentially intransitive. This, however, seems to be a misunderstanding of the distinction between these notions. To avoid this confusion, it is important to stress that the distinction, as well as the distinction between "creature consciousness" and "state consciousness", is a *conceptual* distinction about different uses of the predicate "is conscious" in ordinary language. For instance, in "Pierre is conscious that Mary has just arrived" the predicate "is conscious" is used transitively and in "Pierre's desire to have sex with Mary is conscious" it is used intransitively. But there is no contradiction in saying that when Pierre is conscious that Mary has just arrived (transitively conscious), Pierre is conscious by virtue of being in a mental state which is conscious, that is, intransitively conscious. Moreover, there is no contradiction in the claim that whenever someone is in a mental state which is conscious (intransitive consciousness) the subject is conscious of something (transitive consciousness). One might have theoretical reasons to deny the latter claim, but there is no logical incoherence in that claim. So it is important to be clear that the distinction between transitive and intransitive consciousness is a conceptual distinction between different uses of "is conscious" in ordinary language. For all we know, every instance of a conscious mental state may well lend itself to correct uses of "is conscious" as creature, state, transitive and intransitive consciousness even if, of course, ordinary language hardly allows us to do all that at once. Whether the *nature* of consciousness is to be understood as ultimately intransitive, transitive or creature consciousness is a different issue, and nothing rules out, a priori, that consciousness may ultimately be

necessarily all of that. Of course, proponents of higher-order theories of consciousness notoriously deny that, by pointing out that we can have unconscious states by virtue of which we are aware of something, that is, transitively conscious, like in the well-known example of the inattentive truck driver. If, however, someone denies that there are unconscious mental states, as Brentano does, then it would indeed follow that whenever one is transitively conscious of something one must necessarily be in a mental state which is conscious (intransitively).

What are the implications of this remark for Fisette's discussion? I can see two important implications. The first concerns Fisette's interpretation of Brentano's Thesis II and the second concerns the question whether Brentano's revised theory of consciousness is compatible with the transitivity principle, which I discuss in section 4. As I already pointed out, Fisette presents Brentano's two theses and he advocates an interpretation of Thesis II which conflicts with Rosenthal's (p. 16-17). Let us grant that Thesis I can be interpreted along the lines suggested by Fisette, namely as

1b. Every mental phenomenon is consciousness of something.

Thesis II is as follows:

2. Every mental phenomenon is conscious.

A quite natural reading of 2 is that "is conscious" is used intransitively to attribute consciousness to mental phenomena, or to mental states to use the contemporary terminology. At any rate, from a grammatical point of view, this seems to be the right interpretation. Moreover, this interpretation makes perfect sense in light of the fact that Brentano never accepted the existence of unconscious mental states. Fisette, however, is not satisfied with this straightforward interpretation for the reason that it "stands in contradiction with the first thesis since consciousness cannot be at the same time transitive, as in the first thesis, and intransitive as the second suggests". (p. 16). Why can consciousness not be at once transitive and intransitive, especially if we deny that there exist unconscious mental states as Brentano does? Fisette's reason to reject the straightforward interpretation seems to rest on a misunderstanding of the distinction between the meanings of "transitive" and "intransitive". His claim that it is impossible that consciousness be at once transitive and intransitive would have to be based on some logical or a priori truth. As I have already noted, there is no incoherence in the claim that, on the contrary, consciousness can indeed be at the same time transitive and intransitive. Moreover, this is something that Brentano seems to be committed to, in so far as he denies that there are unconscious mental states. The straightforward intransitive reading of Thesis II is incompatible with Thesis I only if the latter is understood in such a way that it leaves open that one might be transitively conscious of something by virtue of being in an unconscious mental states,

that is, a state which is not intransitively conscious, as proponents of higher-order approaches to consciousness claim. It would seem natural for Brentano, however, to deny that assumption.[1] Thus, Fisette's rejection of the first interpretation of Thesis II is unwarranted.

This being said, it may well be that Brentano's use of "is conscious" in Thesis II is simply ambiguous or indeterminate between the first interpretation and Fisette's proposed interpretation, namely:

2b. Every mental phenomenon is an object of consciousness. (p. 17).

Saying that Thesis II is ambiguous between these two senses seems totally acceptable because, again there is simply no contradiction in claiming that the predicate "is conscious" can be used both transitively to say that a subject is conscious of something and intransitively to talk of a mental state by virtue of which the subject is conscious of something. Finally, it should be pointed out that even if "is conscious" in Thesis II is interpreted as expressing intransitive consciousness, as I have argued, this does not rule out that we may have independent reasons to think that it is plausible to attribute 2b to Brentano. The reason is that given Brentano's claim that every mental phenomenon has both a primary and a secondary object, this strongly suggests that every mental phenomenon is an object of consciousness because every mental phenomenon is a secondary object of consciousness. In other words, there is no incompatibility between Thesis I, the first (intransitive) interpretation of Thesis II, and 2b.

Brentano's revised theory of consciousness and the relation between primary and secondary object

In order to overcome the infinite regress argument, Brentano's strategy consists in denying the third premise according to which "[t]he representation that accompanies the initial mental state is numerically distinct from the targeted state" (p. 19). If so, then insofar as Brentano holds that a mental state has both a primary and a secondary object, we need an account of the relation between the primary and secondary object of mental states, and of how they are unified in "one and the same act". As Fisette notes (p. 23), Brentano holds the following view, as an answer to this question:

3. For any mental state of a subject S, M is conscious iff there is an M and an M**, such that (i) M* is part of M, (ii) M** is part of M, and (iii) M is a whole which M* and M** are part of.*

[1] See Kriegel (2009, p. 28-32) who, in this Brentanian spirit, claims that transitive creature consciousness depends on intransitive state consciousness.

In this statement, M* = representation of the primary object, M** = representation of the secondary object and M = the whole (complex) unifying M* and M**. It is important to recall that 3 stands in sharp contrast to standard higher-order approaches to consciousness such as Rosenthal's HOT theory, because while the latter necessarily involve two numerically distinct states, 3 involves only one mental state. For this reason, standard higher-approaches to consciousness are sometimes called two-state views in contrast to accounts like 3, which are sometimes called one-state views.[2] The question I want to raise is how Brentano's revised theory of consciousness is supposed to account for the relation between the primary and the secondary object, since it is unclear how 3 could play that role.

In section 8, Fisette proposes an interpretation of Brentano's revised theory of consciousness which rests crucially on the notion of a mental agent understood as a "unified real being". The motivation for this notion is that the mental agent is understood as a real substrate, which unifies the various parts ("divisives" in Brentano's terminology) of the mental phenomenon. According to Brentano's revised theory:

> [A] state is conscious only if an agent becomes aware not of this state as such, but rather of himself as being in such a state. Thus, [...] in performing normally, say, an act of external perception the agent becomes aware not only of the primary object, but also of himself as perceiving agent (BRENTANO, 1954, p. 226). This is also confirmed by a passage from the 1911 "Appendix to the Classification of Mental Phenomena" in which Brentano maintains that the object of the secondary consciousness of internal perception is the mental agent himself as constituting both the relationship to the primary object and the secondary consciousness as a relation to the agent himself. (p. 29).

The passage quoted from Brentano states that "the secondary object is not a reference but a mental activity, or, more strictly speaking, the mentally active agent [...] in which the secondary reference is included along with the primary one. (*Psychology*, translation modified, p. 215; *Schfriten* I, p. 385)" (p. 29). It is important to recall that according to Brentano's revised theory, a "unified real being" is something quite peculiar because in contrast to physical phenomena, which have only "intentional existence", a unified real being does not exist only intentionally but it really exist.

How is statement 3 above supposed to give an account of the relation between the primary and secondary object according to Brentano's revised

[2] For the most well-known standard higher-order approaches to consciousness, or two-state views, see Rosenthal (1997, 2005), Lycan (1996, 2004), and Carruthers (2000). For one-state views see, for instance, Kriegel and Williford (2006), Kriegel (2009) and Van Gullick (2006). As Fisette indicates, proponents of two-state views endorse statement 2 (p. 22), as an account of the relation between primary and secondary object, while proponents of one-state views endorse either statement 1 (p. 22) or statement 3 (p. 23). According to Fisette, Brentano's account of that relation corresponds to statement 3.

theory? As I noted, it is unclear that 3 will still work, at least as it stands, because the revised theory rests crucially on the notion of a mental agent, understood as a unified real being, while no mental agent figures as parts of the mental state, or as the whole (complex) mental state in 3.

I can only think of two ways to accommodate the mentally active agent in a revised formulation of 3. Either M, the whole or complex mental state, is the mentally active agent, or M** is the mentally active agent. According to the latter, the mentally active agent, namely M**, would be a part of a whole (or complex), namely M. The mental agent, however, is supposed to be what plays the role of unifying the diverse parts of the mental states. Thus, this option would seem to be a non-starter in so far as it identifies a unified real being with a part of something else. What about the former option? Since M is the whole (or complex) unifying M* and M**, it is tempting to hold that, according to the revised theory, M is the mentally active agent, but then what about M**? We might be tempted to say that M** stands for the agent's activity of representing the secondary object, namely the mentally active agent, that is, M itself. This, however, raises two difficulties. First, M* would have to be the agent's activity of representing the primary object. On such an interpretation, however, M* and M** would turn out to be two different activities: (i) the activity of representing the primary object and (ii) the activity of representing the mentally active agent. Moreover, in what sense would M** be the activity of representing the mentally active agent? What mentally active agent would it represent? Would it represent the agent actively representing the primary object or the agent actively representing herself? Here it might be tempting to say that M** is the activity of representing at once both the primary object and the mentally active agent, namely M. If so, however, M* becomes unnecessary: M is the mentally active agent and M** is a part of M whose function it is to represent both the primary object and the secondary object, namely the whole mentally active agent herself. Thus, this account would be substantially different from 3.

A better way to accommodate 3 within Brentano's revised theory would be to understand M* and M** as two parts of one and the same activity which consists in representing at once both the primary object and the secondary object which, as we saw, is the mentally active agent, namely the whole of M. In what follows, I use this idea to formulate a revised version of the relation between the primary and secondary object.

The second difficulty concerns the nature of the relation, or relational activity, between the mental agent and the primary and secondary object. If Brentano's revised theory should be compatible with 3 along the lines I have just suggested, then the mental activity would have to be a representational activity. If this is so, however, Brentano's revised theory of consciousness would indeed be intentionalist, contrary to Fisette's interpretation. The mental

agent would stand in a representational, and hence intentional, relation both to the primary object and to the secondary object, namely herself, or herself mentally acting. According to such an interpretation, however, the mental agent could no longer be a "unified real being" since it would only be "intentionally existent", like the primary object. These difficulties are serious enough to suggest that attempts to accommodate Brentano's revised theory with statement 3 fail or, at least, that they are far from being straightforward. The second difficulty is a serious problem indeed. Given that the unity of the mental act requires that such a mental act place the mental agent at once in an intentional relation to the primary object and in a non-intentional relation to herself, or to her own mental activity, it seems preferable to give up 3 altogether and to try to find a different account of the relation between the primary and secondary object, that is, an account which is better suited to Brentano's revised theory of consciousness. As a first approximation, the subject S, understood as a "unified real being", actively represents a primary object (say the sound) and is actively in a direct non-intentional relation to herself representing the primary object. More formally:

4. For any mental state M of a subject S, M is conscious iff M is an act of S such that by M-ing S represents a primary object O, and S is non-intentionally, directly aware of herself and of her M-ing.

It should be stressed that, according to this view, the subject is understood as a mental agent in Brentano's sense, namely as a unified real being which does not have only intentional existence. As far as I can tell, statement 4 provides us a clear understanding of the relation between the primary and secondary object, according to Brentano's revised theory of consciousness. In the next section, I assess this view, and point out why I find it implausible and why some alternative views, which are in the spirit of Brentano's theory, should be preferred.

Assessing Brentano's revised theory of consciousness

Before assessing Brentano's view concerning the relation between the primary and secondary object, I should underline an important difference between my interpretation of Brentano's view, namely 4, and Brentano's revised theory of consciousness as it must be strictly understood. This is only a minor point, but an important one nonetheless. Strictly speaking, statement 4 provides a definition of what it is for a mental state to be conscious. This way of talking is, of course, quite relevant in the context of recent discussions in philosophy of mind concerning the nature of consciousness, which try to underscore some essential characteristic of mental states by virtue of which they are conscious rather than unconscious. From Brentano's point of view,

however, it is quite irrelevant to try to give a specific characterization of conscious mental states *per se*, as opposed to unconscious mental states, simply because Brentano never accepted that there exist unconscious mental states. Thus, a more careful formulation of Brentano's account of the relation between the primary and secondary object must consist first in stressing that all mental states are conscious (Thesis II). Whether we interpret Thesis II in the sense of intransitive consciousness or as Fisette's 2b, namely the claim that "every mental phenomenon is an object of consciousness" (p. 17), or even that Thesis II is indeterminate between these two interpretations, it remains that for Brentano it makes no sense to talk of unconscious mental states. Thus, Brentano is a good Cartesian because for him consciousness is the mark of the mental, to use Rosenthal's terminology. A more careful formulation of Brentano's account of the relation between the primary and secondary object, according to his revised theory, would thus be as follows:

4. For any state M of a subject S, M is a mental state of S iff M is conscious, where M is conscious iff M is an act of S such that by M-ing S represents a primary object O, and S is non-intentionally, directly aware of herself and of her M-ing.*

Of course, 4* will strike most contemporary philosophers and cognitive scientists, including myself, as totally implausible and obsolete, in so far as it is highly plausible that there are indeed many unconscious mental states which provide fruitful explanations of psychological phenomena and of human behavior, from an empirical point of view.

This being said, even if we accept that there are plenty of unconscious mental states which can contribute to explain our mental lives, Brentano's revised theory of consciousness can still constitute a plausible account, which is of interest to contemporary philosophy of mind, if his revised theory is understood not as a general account of mentality, but specifically as a theory of what makes mental states conscious, that is, what distinguishes them from unconscious states. This is precisely my interpretation of Brentano's account of the relation between the primary and secondary object, as it is stated in 4. This being said, is that view plausible?

First, this view certainly has the *prima facie* appeal of easily accounting for what Uriah Kriegel calls "the subjective character" of conscious mental states. According to Kriegel's theory, the *bluish* way it is like for me to experience the blue sky has two aspects or components: "the *bluish* component, which I call the experience's qualitative character, and [...] the *for-me* component, which I call the experience's subjective character." (KRIEGEL, 2009, p. 8). Joseph Levine makes a similar point noting that in a conscious experience "there is both a distinctive qualitative character to be reckoned with and *also the fact that the state is conscious – 'for the subject' – in a way that unconscious*

states are not." (LEVINE, 2006, p. 174). If in a conscious experience the subject is non-intentionally, directly aware of herself, understood as a unified real being, then this would readily explain why conscious experiences have a "subjective character".

Secondly, Brentano's revised theory of consciousness would also provide an account of an intuition that some philosophers have recently underscored in the literature.[3] According to this intuition, the awareness one has of one's own conscious states is not an intentional relation and, hence, it cannot be accounted for in the context of a purely representationalist theory of consciousness. According to this line of thought, the relationship between oneself and one's own conscious states is more *intimate* than any representational, or intentional, relationship. As Kriegel (2009, p. 107) points it out, according to this intuition, such a relationship does not involve a gap between the vehicle of representation and the content of representation, while a representational relation does involve such a gap. Claiming that "S is non-intentionally, directly aware of herself and of her M-ing", as it is stated in 4, would be a way to make sense of that intuition.

The first issue that obviously arises, however, is to understand what the non-intentional direct relation of the unified real being to itself is supposed to be. Fisette's presentation of Brentano's revised theory is silent about that. This suggestion, however, strikes me as very similar to Bertrand Russell's view according to which a Self – which is understood in a similar way than Brentano's mental agent, and Descartes's notion of ego for that matter – is in a mental relation of acquaintance to itself. It seems that the best prospect for accounting for this non-intentional relation is indeed russellian acquaintance. Thus, if we have reasons to doubt the plausibility of Russell's notion of acquaintance, this would undermine the plausibility of 4.

More fundamentally, however, why should we accept in our ontology such Brentanian mental agents? Why should we accept that especially if, as Brentano's own view holds it, physical phenomena have *only* intentional existence, while mental agents have a more real kind of existence? This suggests that things can exist in two different ways, which will strike many as very implausible. From an ontological point of view, we may want to say that some entities do ultimately exist while our referential, and denotational, use of language sometimes refer to, or denote, only pseudo-entities, that is things which do not really exists ultimately, but which can be reduced to things that do exist ultimately. Why should we think, however, that mental agents are ultimate existents while physical phenomena are not? Of course, given the plausibility of materialism, we should say that it is the other way around. Moreover, if we can make sense of Brentano's idea that some things exist only

[3] See Levine (2001, 2006) and Hellie (2007).

intentionally, philosophers inclined to accept anti-realism would surely be tempted to say that everything exists only intentionally, and so there is no reason to claim that mental agents exist otherwise. Why should we grant a special ontological status to mental agents?[4] Is it not possible to account for the subjective character, or *for-me-ness*, of conscious experience without making such an ontological commitment? As it is well known, Derek Parfit (1984) has made it clear that we can make perfect sense of our concept of self (or the I-concept) and of agency without accepting substantial selves in our ontology. The general idea of Parfit's reductionist metaphysics of persons is that the self can be reduced to a psycho-physical continuum constituted of mental states which satisfy a certain relation of psychophysical connexion.[5]

This suggests that we may accept Brentano's central claim that every mental state has both a primary and secondary object and also accept, in part, Brentano's understanding of a mental state as a mental activity, while endorsing some different accounts of the relation between the primary and secondary object, that is, accounts which are free of the ontological burden of mental agents, understood as unified real beings. The first alternative account stands in opposition to Brentano's revised theory, in so far as it is obviously representationalist and, hence, intentionalist, and in so far as no mental agent is invoked. An important motivation of that account is that by being representationalist it does not subscribe to Brentano's view according to which the subject and physical phenomena would have different kinds of existence. To put this point in the Brentanian vocabulary, according to this view, the subject would also have *only* intentional existence.

5. For any mental state M of a subject S, M is conscious iff M is an activity of S by virtue of which M represents at once a primary object O and that S's own M-ing is going on.

It is important to stress that while this view if crucially different from Brentano's, it is still brentanian, or neo-brentanian, in the sense that it is also a one-state view, in contrast to the two-state views of standard higher-approaches to consciousness.[6] It should also be noted that, according to this view, there is no need to claim that M is a whole constituted of parts. M is a simple mental activity, which has a complex content, and hence, there is no need to postulate a unified real being. In other words, this view can be understood as a variant of the first account of the relation between primary

[4] See Metzinger (2003) who argues, on the basis of recent empirical findings in cognitive science, that no such things as selves exist, but that there are only "phenomenal selves", namely continuous ongoing processes creating impressions as of a self.
[5] Parfit's metaphysics of persons is very much in the spirit of the Buddhist docrtine of *anatta* (non-self). See Bernier (2011) where I argue for the cogency of the Buddhist doctrine of non-self.
[6] See note 3.

and secondary object discussed by Fisette.[7] We may be tempted to object to 5 on the grounds that it makes no sense to talk of mental activity without postulating a mental agent, understood as a unified real being. This concern can be addressed by pointing out that while it makes no sense to conceive of mental activity without attributing it to a subject, it is an additional and unnecessary claim to identify the subject with a mental agent *qua* unified real being. As long as the activity of M-ing is understood as occurring in a particular mental stream which is constituted of many mental acts which are adequately related to each other, this suffices to conceive the mental act as being the act of a subject. All that is required to account for our intuitions about agency is that the continuum of mental acts exhibits some regular patterns. I pointed out that one of the virtues of Brentano's revised theory is that it allows us to make sense of the idea that conscious states have a subjective character, in Kriegel's sense. This alternative view is also in a good position to account for the subjective character given that M represents that the subject's own M-ing is going on.

The second alternative account is non-intentionalist, but also partly intentionalist. It is non-intentionalist, however, in a different sense than Brentano's revised theory, because it also comes short of postulating a mental agent *qua* unified real being. This account consists in preserving the non-intentional relation – which could perhaps be interpreted as russellian acquaintance – but to interpret it not as a relation to a mental agent *qua* unified real being, but as a relation to the very mental activity itself:

6. For any mental state M of a subject S, M is conscious iff M is an act of subject S which i) represents a primary object O, ii) represents S, as a secondary object and iii) S is non-intentionally, directly aware of M-ing.

This interpretation may seem bizarre, but if we grant that physical phenomena only have intentional existence, as Brentano does, why not grant that the subject also has only intentional existence. Morevoer, if we accept a reductionist metaphysics of persons, along the general lines suggested by Parfit, all we are directly, non-intentionally, aware of is the conscious mental activity; the *relata* of this activity are conceptual constructions, which are represented, as clauses (ii) and (iii) states it. Still, one might insist that clause (iii) is unintelligible unless we postulate a mental agent, in Brentano's sense. According to 6, however, only the mental activity itself is really existent, that is, not only intentionally existent. Moreover, it is not incoherent to add that this

[7] See p. 22 : "1. For any mental state M of a subject S, there is necessarily a mental state M^* such that S is in state M^*, where M^* represents M, and $M^* = M$." Fisette points out that Brentano rejects 1 on the grounds that it entails phenomenalism. It is unclear, however, that 5 entails phenomenalism since it specifies that the mental state must represent a primary object and it is left open that such an object exists independently of any mental activity.

mental activity has a reflexive, or indexical-like, aspect by virtue of which it directly refers to itself[8] and, since the mental act represents the subject as a secondary object, the subject at once thinks of herself as performing this very mental activity which is non-intentionally directly referred to. Another way to put this point in the Parfitian framework is to claim that while the mental activity is a real "non-intentional" existent, clauses (ii) and (iii) indicate how this mental activity locates itself in a particular mental continuum. Hence, no mental agent *qua* unified real being is required to make sense of clause (iii).[9]

Just as interpretation 5 is able to account for the subjective character, this interpretation (6) is also able to do so. There is a difference with 5, however, that some may find appealing. I pointed out that, by invoking a non-intentional relation to the mentally active agent, Brentano's revised theory is in a good position to account for the intuition that the relationship a subject has to her own consciousness is more *intimate* than any representational relation. In contrast to 5, interpretation 6 is in a good position to account for this intuition of intimacy, as long as it is able to make sense of the non-intentional relation, perhaps in terms of Russellian acquaintance, as I suggested.

There is also an alternative intentionalist view which combines intuitions of both 5 and 6. It consists in claiming that the representational content "that S's own M-ing is going on" in 5 is a *de re* proposition, in the sense that the very M-ing is partly constitutive of its own representational content. Thus, this variant preserves the idea that the mental activity itself is part of the content of the mental state, which we find in 6, but in contrast to 6, the content is fully representational. Since this view is representationalist, we can call it 5*. This interpretation is very appealing. It preserves the idea that the mental activity is directly referred to, which we have in 6. This is plausible because intuitively a conscious mental state seems to have a kind of inner presence to itself, which does not require conceptualization. In other words, this interpretation makes representationalism compatible with the intuition I noted above concerning the intimate character of the relationship between one's awareness of one's conscious experience and the conscious experience. Moreover, if we accept the parfitian view of persons, the direct reference to this mental activity provides an anchor to the mental continuum which is constitutive of the

[8] See Bernier (2010 and 2011) where I have proposed a view along these lines.

[9] If we accept Brentano's distinction between intentional existence and real existence, this interpretation, entails a form of phenomenalism, because only the mental activity would have real existence, as opposed to intentional existence. If, however, we deny Brentano's claim that physical phenomena have only intentional existence, and if we accept that the mental activity is actually some neurobiological activity going on in the brain, then this interpretation would not entail phenomenalism. From an ontological point of view, it would be compatible with realism. According to such a realist interpretation, the distinction between the primary and secondary object, on the one hand, and the mental activity itself, on the other hand, would not be an ontological distinction but only an epistemic distinction. While the subject has an epistemic access to the primary object and to herself, only via some representation, she has a direct epistemic access to her own mental activity.

reductive basis for the mental subject. Finally, since there is no incoherence in thinking that such a reductive basis might eventually be accounted for in terms of a neurobiological continuum, such a *de re* proposition might ultimately be directly referring to some occurring brain activity. Compared to 6, however, 5* has the strong advantage of not invoking a controversial non-intentional relation. As I already pointed out, the best prospect for accounting for that non-intentional relation is to characterize it in terms of russellian acquaintance. The appeal of 5* is that it is free of this controversial theoretical burden.

Moreover, as I pointed out, 6 is committed to the claim that the mental activity which is in a direct non-intentional relation to itself has some kind of ontological priority and thus it can easily lead to phenomenalism. This is not the case with 5*, however. Since the content of the mental state is fully representational, it does not need to invoke Brentano's claim that there is an ontological asymmetry between things that exists only intentionally and those, such as mental agents, which have more than intentional existence. According to 5*, the objects of conscious mental states all have the same ontological status. Whether we want to say that they all exist only intentionally, as anti-realism would have it, or that they exist in a more robust sense, as realists would have it, is a further independent issue on which 5* remains neutral.

Since 5 is compatible with 5*, it is useful to distinguish it from a more determinate interpretation of 5, which is incompatible with 5*. According to this interpretation of 5, which we may call 5', the propositional content "that this very M-ing is going on" in 5 is not a *de re* proposition. In addition to being fully representationalist, 5' could be understood also as conceptualist in the sense that all the elements which constitute the representational content are conceptual.

Compared to interpretations 5*, 6 and 5', Brentano's revised theory of consciousness, which I have characterized as claim 4, seems very implausible because it rests on the dubious notion of a mental agent *qua* unified real being which, by definition, cannot be something physical. Moreover, this view presupposes a dubious ontological distinction between the ontological status of such a mental agent and things that have only intentional existence, suggesting that mental agents are, so to speak, more real than any physical phenomena. As I have indicated, this view is unacceptable both from a realist point of view which accepts physical phenomena in its ontology and from an anti-realist point of view which denies the existence of brentanian mental agents, not to mention Cartesian egos. While Brentano's revised theory of consciousness is interesting in suggesting a way to account for the subjective character, or *for-me-ness*, of conscious experience, it rests on an unnecessary ontological distinction between things which have only intentional existence and things which have more existence than that. The appeal of fully representationalist views, such as 5* and 5', is that they do not require such a

distinction. Moreover, as I have argued, while we can accept that conscious mental states do necessarily involve a subjective character, as many philosophers have recently proposed, it is not necessary to postulate the existence of a mental agent in Brentano's strong ontological sense, in order to account for the subjective character. In light of a reductionist metaphysics of persons, à la Parfit, the views suggested by 5* and 6 can also account for subjective character. In so far as 5* and 6 would seem to presuppose such a reductionist metaphysics of persons, however, they must postulate the existence of mental conscious states which are inscribed within some psycho-physical *continuum*. This is hardly a problem, however. Statement 5', on the other hand, lends itself to a more radically anti-realist view, because it is not committed to the existence of such a psycho-physical *continuum*. On such a radical view, conscious mental states would be just as much conceptual constructions, as the primary object and as the subject. While 5* is the view I find most appealing, for the reasons I have suggested, and 6 seems problematic, given that it requires a dubious relation of acquaintance, I must admit that I do not find 5' implausible. Be that as it may, one thing is clear: Brentano's revised theory of consciousness is indeed implausible and 5*, 6, or 5' should be preferred.

Brentano's revised theory of consciousness and the transitivity principle

Fisette's paper makes much of comparing and contrasting Brentano's theory with higher-order approaches to consciousness and, especially, with Rosenthal's HOT theory. As Fisette recalls, the latter rests fundamentally on the transitivity principle according to which a mental state is conscious if, and only if, one is conscious of that state (ROSENTHAL, 2005, p. 179). The question I want to raise is whether Brentano's revised theory of consciousness is compatible with the transitivity principle. Fisette's position seems a bit unstable, in that respect. Towards the end of the paper he notes: "Brentano's theory of consciousness is not consistent with the principle of transitivity" (p. 32). This statement is puzzling because not much earlier in the paper he states: "this new version of Brentano's theory of consciousness is not incompatible with Rosenthal's transitivity principle". (p. 30). What are we to make of these claims, which seem contradictory? My formulation of Brentano's view, as statement 4 above, may help. Statement 4 makes it clear that, according to Brentano's revised theory of consciousness, when a subject has a conscious mental state, she must be non-intentionally, directly aware of herself and of her mental act. If we suppose that the transitivity principle entails that "one can be conscious of one's mental state" only if one is in an intentional relation to one's mental state then, of course, Brentano's revised theory of consciousness would indeed be incompatible with the transitivity principle. Why, however,

should we accept that the transitivity principle has this intentionalist implication? After all, even if the subject is *non-intentionally* directly aware of herself and of her mental act, she is still aware *of something*. The subject is in a transitive or "objectual" relation to her own mental activity, which after all may only be some processes going on in her brain.

Moreover, as I argued in section 1, even if we reject Fisette's interpretation of Brentano's Thesis II as "2b. Every mental phenomenon is an object of consciousness", it is still plausible to attribute 2b to Brentano. Saying that every mental phenomenon is an object of consciousness, however, comes very close to the transitivity principle. How can a mental phenomenon be an object of consciousness otherwise than by the subject being aware of that phenomenon? It is not as if 2b should be understood as claiming that a third person is aware of the subject's mental state. The upshot is that Brentano's revised theory of consciousness is compatible with the transitivity principle. As I pointed out, however, this should not be understood as entailing that Brentano's theory is intentionalist as my statement of the revised theory makes it clear.

There may be reasons to object to Fisette's interpretation of Brentano's revised theory of consciousness on exegetical grounds. In this paper, however, I have taken Fisette's interpretation at face value. I have argued that statement 4 captures the gist of Brentano's revised theory of consciousness, as Fisette interprets it. I have pointed out that while this theory has the virtue of accounting for the properly subjective character of conscious mental states, or their *for-me-ness*, and of accounting for the intuition of intimacy, still it carries an unnecessary ontological burden by postulating the existence of a mental agent understood as a unified real being. I have underscored some variants of Brentano's theory which are free of this ontological burden and I have argued that these variants should be preferred to Brentano's own theory. To conclude, it is important to stress that while the variants of Brentano's theory (namely, 5*, 5' and 6) are substantially different from Brentano's theory, these views can still be called Brentanian, or neo-Brentanian, in the important sense that they all correspond to what has been called "one-state views", in the literature. As Fisette makes it clear, two-state views are denied by Brentano in his reply to the infinite regress objection.

In so far as Brentano's revised theory and the three alternative views (5*, 5' and 6) are one-state views, does this mean that according to all these views, consciousness is an intrinsic property of conscious mental states? I will not attempt to give a definite answer to this question here, but I only want to make the obvious point that this depends on what we mean by an "intrinsic property". It seems plausible to understand Brentano's revised theory as entailing that consciousness is indeed an intrinsic property of mental states, as Fisette claims, because Brentano's theory is in agreement with the Cartesian view according to which consciousness is the mark of the mental and because his

revised theory rests on the postulate of a mental agent understood as a unified real being. In this sense, consciousness turns out to be something *sui generis* which is irreducible to anything else. Thus, it seems safe to say that according to Brentano's revised theory, consciousness turns out to be an intrinsic property in a fairly strong sense. It remains unclear, however, whether according to variants 5*, 5' and 6, consciousness is an intrinsic property. It is certainly not intrinsic in the sense that it would be the mark of the mental, where "the mental" is understood as something necessarily non-physical. If the claim that consciousness is an intrinsic property of mental states is understood in the sense that it is *sui generis* and irreducible to anything else, then it is plausible that according to these variants, consciousness is not an intrinsic property, because nothing rules out a priori that these views be compatible with physicalism.

References

BERNIER, P. "La pensée sans sujet pensant". *Dialogue*, v. 49, p. 1-14, 2010.

_____."Is the Buddhist Doctrine of Non-Self Conceptually Incoherent?", *Buddhist Studies Review*, v. 28, p. 187-202, 2011.

BRENTANO, F. *Psychology from an Empirical Standpoint*, translation by A. C. Rancurello, D. B. Terrell and L. L. McAlister, London: Routledge and Kegan Paul, 1973.

CARRUTHERS, P. *Phenomenal Consciousness. A Naturalistic Theory*. Cambridge: Cambridge University Press, 2000.

HELLIE, B. "Higher-Order Intentionality and Higher-Order Acquaintance", *Philosophical Studies*, v. 134, p. 289-324, 2007.

KRIEGEL. U. *Subjective Consciousness. A Self-Representational Theory*, Oxford: Oxford University Press, 2009.

_____. and K. WILLIFORD. *Self-Representational Approaches to Consciousness*. Cambridge, MA: MIT Press, 2006.

LEVINE, J. *Purple Haze. The Puzzle of Consciousness*. Oxford: Oxford University Press, 2001.

LEVINE, J. "Conscious Awareness and (Self) Representation". In: KRIEGEL, U. and WILLIFORD, K. (Eds.). *Self-Representational Approaches to Consciousness*, Cambridge, MA: MIT Press, 2006, p. 173-197.

LYCAN, W. G. *Consciousness and Experience*. Cambridge, MA: MIT Press, 1996.

_____. "The Superiority of HOP to HOT". In: GENNARO, R. (Ed.). *Higher-Order Theories of Consciousness*. Amsterdam: John Benjamins, 2004, p. 93-114.

METZINGER, T. *Being No One. The Self-Model of Subjectivity*. Cambridge, MA: MIT Press, 2003.

PARFIT, D. *Reasons and Persons*. Oxford: Oxford University Press, 1984.

ROSENTHAL, D. M. "Two Concepts of Consciousness". *Philosophical Studies*, v. 94, p. 329-359, 1986.

_____. "A Theory of Consciousness". In: BLOCK, N.; FLANAGAN, O. and GÜZELDERE, G. (Eds.). *The Nature of Consciousness. Philosophical Debates*, Cambridge, MA: MIT Press, p. 729-753, 1997.

_____. "Sensory Qualities, Consciousness and Perception". In: ROSENTHAL, D. M.: *Consciousness and Mind*. Oxford: Clarendon Press, 2005. p. 175-226.

TUGENDTHAT, E. *Self-Consciousness and Self-Determination*, translation by P. Stern. Cambridge, MA: MIT Press, 1979.

VAN GULICK, R. "Mirror Mirror – Is that all?". In: KRIEGEL, U. and WILLIFORD, K. (Eds.). *Self-Representational Approaches to Consciousness*. Cambridge, MA: MIT Press, 2006, p. 11-39.

ARGUMENTOS
Revista de Filosofia

Pedro M. S. Alves*

Comments on Fisette's: "Franz Brentano and higher-order theories of consciousness"

ABSTRACT

In this paper, I discuss Fisette's Interpretation of Brentano's philosophy of mind in the framework of the modern higher-order theories, especially Rosenthal's version. While acknowledging the truth of Fisette's rendering of Brentano's psychology as a first-order theory, I examine at length the theses that all mental states are conscious states, or can became conscious states, and that the first-order consciousness of some mental states must be accounted for as self-consciousness. I disagree with both theses, and I propose a general picture of mental states and consciousness in keeping with some insights coming from Leibniz, and not from Aristotle or Descartes.

Keywords: Philosophy of mind; Brentano; Psychology; Higher order theory of consciousness; Self-consciousness; Leibniz.

RESUMO

Neste artigo, discuto a interpretação de Fisette a respeito da filosofia da mente de Brentano contra o pano de fundo das modernas teorias de ordem superior, especialmente na versão de Rosenthal. Ao reconhecer a verdade da interpretação de Fisette da psicologia de Brentano como uma teoria de primeira ordem, examino detidamente as teses de que todos os estados mentis são estados conscientes, ou podem se tornar estados conscientes, e de que a consciência de primeira ordem de alguns estados mentais deve ser explicada como auto-consciência. Discordo com ambas as teses e proponho uma imagem geral dos estados mentais e da consciência a partir de algumas intuições vindas de Leibniz e não de Aristóteles e Descartes.

Palavras-chave: Filosofia da mente; Brentano; Psicologia; Teoria de ordem superior da consciência; Auto-consciência; Leibniz.

* Universidade de Lisboa. E-mail: psalves2@gmail.com

In his sound, interesting and perspicuous paper, Professor Denis Fisette addresses several important issues, namely:

1. The correct understanding of Brentano's theory of self-consciousness;
2. Brentano's place in the framework of modern higher-order theories; (HOTs) of (self-) consciousness, especially Rosenthal's;
3. A vindication of Brentano's theory of intransitive and intrinsic self-consciousness, reinstated in the wider context of his latter writings;
4. The overall connection between mental states and consciousness.

The general discussion of these issues involves, as a background, some particular conception about the highly controversial relationship between, on the one hand, consciousness and mind, and, on the other hand, between the (conscious) mind and the physical system where mental states and conscious states do occur. Indeed, to account for the relationship between conscious states and mental states, and eventually between the mind and the physical system underlying it (a question that is left aside here) is a task that is not independent from the theoretical position one wants to uphold regarding the definition of what mental and conscious states by themselves are.

Indeed, can we simply equate mental states and conscious states, so that there are only conscious mental states and, therefore, an overlap between consciousness and mind? Or, the other way around, is there room for an in principle distinction between mental states and consciousness, so that either consciousness is reducible to some relational feature of mental states (for instance, overall access), or it becomes a property superimposed on some (but not all) mental states?

Lurking in the twilight, and twinkling here and there in the paper, there is, in addition, the paramount question whether it still remains an acceptable image of our mental life the common idea – which has its roots in Descartes as well as in Locke – of a conscious self that knows everything that happens on his own mind, and which is in control of all that goes on in his mind because he is conscious of it. In short: are the ideas of a unitary (and unique) stream of consciousness, and of a non-analyzable conscious subject piloting (or at least accompanying) his own mental life something we still can be comfortable with?

If this Cartesian and Lockean view shows itself as untenable, there are other models available on the market besides an "Aristotelian" conception of the mind. Indeed, there is something of a false dilemma in the way Rosenthal invites us to choose between an Aristotelian-like or a Cartesian-like conception of mind and consciousness. In modern ages, certainly, the Cartesian approach ruled supreme. However, another alternative powerful conception of mind was put forth in modern ages by Leibniz. His emphasis on the "*petites perceptions*"

and on the "*perceptions inaperçues*" involves a clear separation between representational states (perceptions) and self-consciousness (apperception), so that, as a consequence, there are in the mind, at any time, plenty of perceptions that remain unconscious. This emphasizes the deep complexity and richness of our mental life, in opposition to the very few perceptions that, one after the other in the stream of our reflective life, reach the level of consciousness. Additionally, the separation made by Leibniz implies not only that there are perceptions that are in fact without self-consciousness, but also the much stronger thesis that there are perceptions in the mind that will remain forever unnoticed to self-consciousness. Moreover, apperception is a matter of degree – definitely, there is not a yes-no question regarding it; apperception decreases to zero at a certain (variable) limit – nevertheless, the perceptual life of the mind still continues below this limit. And what the unconscious perceptions carry with themselves is not a conscious-self, but instead the unity (the coherence) of a unique point-of-view rooted in the entire organic system, so that, when conscious life arises over this complex unity of mental life, the reflective grasp of one's I rests on this "mute" unity of a point-of-view and allows therefore a new kind of mental life, where new things arise, like decision-making and reflective thinking, things that make up the mental life of a human person.

Are we, thus, constrained to choose between consciousness and intentionality as the essential mark of the mental? No. There is a deeper vantage point rooted in the unity of the entire organic system, in its mute construal of a point-of-view which encompasses, in a systemic unity, physical processes (the so-called "physical basis" of mental life), perception (intentionality), and apperception (self-consciousness).

My considerations on the engaging paper by Professor Denis Fisette are to a certain extent tributary of this Leibnizian insight about mental life. This is a "*parti pris*" I must state at the outset.

What Brentano says?

Professor Fisette argues convincingly that Brentano's theory of mental-state consciousness is in keeping with an one-level account of consciousness, i.e., that the mental state by which an organism (having "creature consciousness") is transitively conscious of something is itself, at the same time, an intransitively conscious state. The content of this intransitive consciousness is *self*-consciousness. Thus the formula: a mental state is transitively conscious of something and intransitively conscious of itself, or rather, in Brentano's own words as phrased by Professor Fisette, (i) every mental phenomenon is a consciousness (*Bewusstsein*), and, simultaneously, (ii) every mental phenomenon is conscious (*bewusst*), so that, we could add, there is no mental phenomenon directed to something as an object (say, a

physical phenomenon) which does not represent its own occurrence as a mental phenomenon.

However, there are some weird things on these formulae.

Firstly, a kind of "diplopia", inasmuch as every mental state is accounted for as displaying two representational contents and having, thus, two objects: the intentional object as such (as its primary object) and itself (as a secondary object).

Secondly, a somewhat baffling displacement of the expected locus of self-consciousness: indeed, in what sense can we say that a mental state is conscious of *itself,* rather than that there is a consciousness of the self through the mental state?

Finally, if Fisette's exegesis is accurate (and it seems to me that it is), there is here a clear commitment to the controversial thesis that all mental states are by themselves self-conscious states, and, then, that self-consciousness, understood as an intrinsic, non-relational property, is an essential element of mental states as such. This last exegetic assumption contravenes the long-established – and newly reinstated in the contemporary philosophy of mind– interpretation of Brentano's psychology, provided that it puts intransitive self-consciousness, and not only intentionality, as a fundamental feature of the mental. But Fissete's defense of this point against "intentionalism", based as it is on an attentive analysis of the relevant passages of Chapter II of *Psychology,* looks like a comprehensive reading of Brentano's global position.

However, as I said, all this strikes me as problematic.

To begin with the third point, I think that a disambiguation of the consciousness-thesis is required. Is Brentano endorsing the thesis that

A. All mental states are conscious, or rather the thesis according to which
B. We have consciousness of all our mental states?

The second version, as stated in B., is pretty compatible with Rosenthal's position, provided that he would be willing to acknowledge that a higher-order thought targeting a lower-order thought is always and everywhere possible. Despite the fact that, in our mental life, not every first-order thought gets to be a conscious mental state by means of a second-order thought, it would be conceivable that some other more powerful mind than ours will target all its mental states by second, third order thoughts, achieving not only consciousness of all its mental states, but also reflective introspection about all its mental life. What prompts a HOT, according to Rosenthal? Is it the simple existence of an unconscious mental state? Is the bare existence of a mental state a sufficient condition for a suitable HOT? If this is so –and I am not sure if it is– then Brentano and Rosenthal would not be in a radical disagreement.

It appears obvious that Professor Fisette is trying to present Brentano's consciousness-thesis in a quiet different sense. It is not the case that an unconscious mental state prompts by itself the (possible) occurrence of a suitable HOT; rather, the case is that unconscious mental states do not exist at all. Thus, for Professor Fisette, the version A. is the good one. Being so, self-consciousness appears as the fundamental mark of mental life, because, if self-consciousness did not exist, then consciousness of an intentional content would not exist either. Certainly, the converse assertion is also true: if an intentional act did not exist, then self-consciousness would not exist too. Nevertheless, the intentional relationship to an object seems, in this construal, to be a basis, a requirement, for a total act that has in self-consciousness its own achievement. Tracing this idea to its final reinstatements in Brentano's latter writings, Professor Fisette comes eventually to acknowledging that all consciousness is *de se* (we are talking, of course, about intentional states directed to primary objects), containing, thus, not only an object, a sound heard, say, but an "implicit" self-awareness of the very subject which is in the psychic activity of hearing the sound. It seems that performing this loop-like self-awareness is the ultimate end of mental life, as if the self would be able to know everything that goes on in him, and thus to gain control over his entire life. If thesis A. is the accurate interpretation of Brentano's position, then we find here a point of divergence between him and higher-order theories.

For reasons that I shall explain in a while, it seems to me that theses A. and B. are both incorrect. This is no more an exegetic issue. Despite the fact that Professor Fisette is right in ascribing to Brentano the position expressed in A., the question whether thesis A. is right still remains, and I am not full convinced by the arguments Fisette puts forward. I will argue this point later. For now, I want to address the second weird aspect of the formulae above.

As a matter of fact, I cannot find a good sense for the assertion that a mental state is conscious of (or for)… *itself!* A mental state has a representative content. By means of it, the mental state intentionally refers to an object, either physical or mental. In addition, does the mental state have a content that represents the very mental state, namely its own occurrence? I ask: represents for whom? Can we say: for him (or it)? Is this the answer? But a mental state is not the subject of mental life; it is an event in the life of a mind. And what means to represent? This representation must not be construed as an intentional relation to an object which is transcendent to the act itself. This is all the point with the idea of an intrinsic, intransitive consciousness. But how would be *expressed* the sense of this intrinsic, intransitive self-consciousness (if there is any)? If the mental state is accounted for as a close unity referring intrinsically to itself, the expression of this self-representation would necessarily be: (i) "there is a sound" (intentional, primary object), and (ii) "there is a hearing of a sound" (secondary object). However, at least for us, the normal expression of

the last content is "I hear a sound", or "I am hearing a sound", or, in order to grant to the opponent everything we can, "there is a hearing of a sound, and that hearing is mine". But the inclusion of one's I, which is in the state of hearing that sound, amounts to introducing a more complex content that cannot be shrunken in the mental state itself. In a word, if there is something like an intransitive consciousness, and if this intransitive consciousness can be accounted for as a self-consciousness, then we must say that it is not the mental state that is conscious of itself as a secondary object, but rather that a subject is conscious of being himself in that mental state.

In the interesting last section of his paper, Professor Fisette addresses this issue, acknowledging "the ambiguous status in *Psychology* of the concomitant consciousness that accompanies all mental states", and that "a state as such cannot be said to be conscious (or not)". The revisions introduced by Brentano in his last writings consist in introducing the notion of a psychic agent as the subject of mental life, and in making a distinction between expressly noticing (*bemerken*) some content or having just an implicit consciousness of it. This last move bestows some plausibility to Brentano's account. A state can be said to be (intransitively) "conscious" only if an agent becomes aware of himself as being in such a state. This self-consciousness, insofar as the intentional act is performed by a psychic agent, is merely implicit in the very act being performed, so that, as Professor Fisette avows, this implicit (self-) consciousness accompanying the act could be described as a pre-reflective consciousness of the agent itself, when performing an intentional act and before reflectively taking notice of his own psychic activity.

Hence, I wonder if, in the light of this last reappraisal by Brentano himself, we can continue sustaining that the mental act is "conscious" in an irreducibly intrinsic and intransitive sense. To begin with, this consciousness does not belong to the act itself; it is, rather, the consciousness that a subject has of being himself in a determinate mental state. Secondly, this last consciousness of the subject's own activity when performing a mental act is declared to be only "implicitly" present in the mental state as such. Therefore, the following conclusion seems to me unavoidable: to say that a mental act is implicitly conscious is only a name to describe the capability of the act to be subsequently apprehended by a subject as its own mental state. Phrasing this in Rosenthal's jargon, we have: a mental state is (intransitively) conscious if and only if it can be (transitively) conscious for another mental state. Call this capability an "implicit feature" of every mental state and name it "pre-reflective consciousness", if you want. The question is that, in this new light, Brentano's position is not incompatible with Rosenthal's explanation of self-consciousness, and, even more, it is virtually reducible to it. To insist that the mental act was already (intransitively) "conscious" is just a way of circumventing the huge problem of describing how and why a mental state can be captured by a subject

as its *own* mental state (the so-called first-person privileged access). If an extrinsic remark is here allowed to me, I would suggest that Husserl's distinction between the pre-phenomenal being of lived-experiences *(das präphänomenale Sein der Erlebnisse*) before reflection, and their being as phenomena, when reflective turning-to (*Zuwendung*) constitutes them as objects (see Hua X 129), is a possible way-out for the difficulties affecting Brentano's theory: mental states *are not* phenomena before reflection constitutes them as such, and mental-living (*erleben*) *is not* intentionally seizing an object.

Nonetheless, let me now return to *Psychology,* in order to address the first weird feature I pointed above in the formulae. I will be brief, because this is a well-known criticism directed to Brentano's theory of primary and secondary objects. As the mental act is presented as having two representational contents, the second being the side-representation of the act itself (as secondary object), this second act must be itself represented by a third act, and so on. Professor Fissete tries to contravene this line of reasoning pointing to Brentano's mereological distinction between wholes formed by collective and by divisive parts. While collectives can be analyzed in parts that are mutually independent, there are wholes whose parts are simple *abstracta* that have no independent existence outside the whole. Such is the case with the representation of the primary object and the representation of the secondary object in the mental phenomenon containing both. Therefore, the objection seems to be blocked, because we can no more say "as the side-representation is an *act,* so…" I think, however, that the issue is not settled with this move. My point is that the new presentation of Brentano's thesis states that every intentional relationship to a primary object must contain the consciousness of the entire act as its secondary object. However, this second part, representing the total act (and, thus, once more the primary object), is itself a part that must be conscious by a tertiary part, and so on.[1] All in all, there is no clear cutting line between Brentano and Rosenthal: the mental acts have a double representational content, and the second act (or the part) is in both theories accounted for as a case of self-consciousness, so that the real disagreement is limited to the question whether this self-consciousness is a second act or a divisive part of the former act. In my opinion, this is the reason why Brentano's position in the circle of HOT theorists oscillates between the partial acceptance and the partial refusal. We could say that, if HOTs are reducible to a brentanian-like one-level account, this is the right way to go, because a one-level theory is more economic and simple. However, there is a huge obstacle to embark in such a kind of reduction. For a higher-order theory, the accomplishment of lower-order acts does

[1] Let me try to express this progressive growth of elements inside the mental act: (i) there is a sound, (ii) there is the hearing of the sound, (iii) I am conscious of the sound heard and of the hearing of the sound, (iv) I know that I am conscious of the sound heard and of the hearing of the sound, (v) I am conscious that I know that I am conscious of the of the sound heard and of the hearing of the sound, etc.

not depend on the accomplishment the higher-order acts. A first-order thought, a FOT, say, is not dependent on the existence of a SOT or a TOT. We have a progression to infinity which remains merely potential: for a mental state to be performed there is no need of a second mental state targeting the first, and for that last one to be performed, there is no need of a third order one targeting it, while, as a matter of principle, this progress to ever new strata always exists as an open possibility. Regarding Brentano's account, the situation is, however, completely different. We bump here into a *regressus in infinitum* inside the very first-order mental act. In such a case, the mental act could not be performed at all, given that it would contain an infinite number of regressive internal conditions that could never be satisfied. Clearly, the infinite progress and the infinite regress are pretty different. The first is only potential and does not appear as a condition of the lower acts; the second is an actual one that prevents the accomplishment of the very first order act. Surely, our mental life has no such a type of (bad) complexity.

What Rosenthal does not say?

Rosenthal has a point, although he does not wind up the debate with it.

His point is: there are non-conscious mental states. My surmise is that the great majority of our mental life is constituted by such mental states. Indeed, should we, as living organisms, negotiate our transactions with the surrounding world with the few mental states of which we are conscious, dispensing with the mental routines that run its course unnoticed, then we would have disappeared as a living species a long time ago. After all, Brentano's definition of a mental state is question-begging. Mental states are those of which we are conscious and, if they are not, so they are not mental states. This amounts to a dramatic impoverishment of our mental life and to a limitation to the first-person access. The states about which we do not have a first-person access are not conscious for us, of course, but this is not tantamount to saying that they are not mental states at all. Clearly, we are in need of a non-question-begging definition of what a mental state is.

Armstrong's well-known example of the inattentive truck driver shows that, while having creature consciousness, one can be conscious of something without being in a conscious state. Certainly, everything that, in the example, the inattentive driver was not conscious during his trip (the red light he saw, the gear changes he done, the slowdowns he has done at the curves in the road, the proprioception he had of his body pulling to the opposite side of the curves, etc.) was something he *could* have been conscious. Moreover, it is by reference to the conscious mental states he has (seeing a red light, etc.) that he can afterwards "guess" and categorize the mental states he was not conscious of ("I must have seen the red light", etc.) This apparently restores the supremacy

of mental state consciousness. By the same token, absorbed in writing, I suddenly pay attention to an irritating noise, and I realized that he had lasted for a while, and that the nervous way I was shaking my right leg was a consequence of it. The first conclusion is that I must have heard the noise long before my state consciousness of it, so that there was a mental state (a particular sensation) that was running its entire course non-consciously. However, the second conclusion contravenes the first: only after the conscious mental state, and by reference to it, was I in conditions to talk about my previous hearing of a noise.

What is the lesson of these two contradictory trends? One possible answer amounts to saying that the hearing I was talking about was not a real hearing, but only a blend of physical phenomena pertaining to neurology and human motricity. We would be, thus, completely within the Cartesian split between matter and mind. In the very Cartesian formulation, the soul feels "*par occasion de*" certain physical phenomena occurring in the body, but these phenomena are not sensations until they become conscious: sensation is an actual, qualitative, conscious state of the mind; the underlying phenomena are not mental states in the pregnant sense. So I was hearing nothing; I was not hearing at all.

This is an odd conclusion. The opposite lesson is much more fruitful for a definition of a mental state and, what is more, for a definition of *consciousness.* First of all, it must be admitted that there are non-conscious mental states, if one intends to go beyond the old (and odd) Cartesian divide. Secondly, only after this move can we get a productive characterization of what a mental state is, paying attention to how it associates and blends with the overall functioning of the brain and the whole body, so that there is not a dividing line between neurological and mental phenomena, but only a functional definition of which part of the whole can be categorized *as* a mental state. Thirdly, only when we would arrive at a definition of mental states disregarding consciousness would we be in conditions to productively ask about what consciousness *is,* as a new dimension superimposed on some of them. However, we always fix the several types of mental states by reference to the conscious ones. This is true, certainly. Nevertheless, the lesson is that we must know these mental states, have a direct, first-person acquaintance in order to talk about them. But we also must be acquainted with insects in order to do taxonomical entomology; nevertheless, our acquaintance with them is not a part of the taxonomy we are making. I believe that the same holds for mental sates.

Perhaps there is a third approach based on the Brentanian principle of the unity of consciousness, which Professor Fisette emphasizes in order to deal with the problems regarding the relationship between primary and secondary objects. As Professor Fisette writes, quoting directly texts from *Psychology,* "the totality of our mental life, as complex as it may be, always forms a real

unity – this is the well-known fact of the unity of consciousness". Thereby, mental phenomena are singled-out as "partial phenomena of one single phenomenon in which they are contained as one single and unified thing". Perhaps this containment includes, in a total consciousness, at any time, the entire set of mental states we have, in such a way that they are not unconscious mental states, but, instead, states fused and blended together in a global consciousness. This is tantamount to saying that there is, at any time, a unique consciousness, with a focal attention detaching some features, while a peripheral attention gathers the others in a fuzzy way. But I cannot see how sensations, perceptions, thoughts, beliefs, and so on, could be blended in a total self-consciousness. The blatant heterogeneity they exhibit prevents their inclusion, as partial-phenomena, in a putative "single phenomenon" containing them all. Besides that, to return to Armstrong's driver, he *guesses* that he must have seen the red lights in the road: he cannot pick-up retrospectively, by analysis, these perceptions from a supposed global mental state he would remember. Against this hypothesis, my conviction is that the mind has, at any time, a plenty of mental states, which run their course in a parallel organization, without any monitoring center that could cover the entire complexity. Notwithstanding, it would be an odd thing to deny that there is such a monitoring center bringing, at any time, to consciousness *some* of the mental states that occur in the mind.

In fact, there is a clear phenomenological difference between non-conscious and conscious mental states. This difference is not explainable in Rosenthal's framework, as a difference between a mental state transitively conscious of something and the fact that this mental state comes to be targeted as an object by another mental state. In a word, as Brentano saw and Professor Fisette rightly underlines, a conscious mental state includes something in it that is irreducible either to the transitive consciousness inside the first order-state or to reflective seizing by a higher-order thought.

To put the things in order, let us imagine a mental life such as ours which, as a matter of fact, would never perform higher-order thoughts targeting its first-order mental states. If we believe in Rosenthal's account, this mind would be like a zombie or a robot, or it would be like the inattentive truck driver for *all* its mental states. But is this a reasonable hypothesis? However, it seems to me that it is unavoidable according to Rosenthal's explanations. Against the inattentive driver example, I now propose to considerer the example of the attentive listener. Suppose someone listening to a symphony in an audience hall; and suppose she is totally immersed, absorbed, in such a way that the perception of the surroundings and of her own body disappear almost completely from the focus of her attention, even if these mental states still continue in a "truck driver's" way. In a word, for her, in such a state of self-forgetfulness, there are only these sounds that she listens to, and no higher-

order thought about them breaks the ecstatic experience she is living. Now, clearly she has perceptual states about the surroundings (the chairs, other people in the hall, etc.), and perceptual states about the music she listens to. All they are first-order mental states. Nevertheless, there is a notorious difference between the first-order perceptions of the symphony and the first-order perceptions of the other things in the audience hall. This difference, *internal* to the first-order mental states, is the difference between conscious and non-conscious mental states.

I agree, thus, with Professor Fisette that this kind of consciousness is intrinsic. But I disagree with Brentano and Fisette when they construe it as a case of *self*-consciousness. I think that, before self-consciousness, there is a difference in the very *mode of giveness* of the objects. What is, then, consciousness, as a mark of some mental states? Here, we have only intuitions, when it comes to propose a non-circular characterization, as it is the case when, for instance, we talk about "awareness" for explaining "consciousness". There are several proposals on the market: phenomenal-consciousness, what it is like, worldly versus experiential subjectivity, thin versus thick phenomenality, and so on. So let me express my intuition too. It is based on a strategic move. Why not asking the truck driver himself about what he feels bizarre when, astonished, he realizes that he has reached his destination?

First bizarre aspect: time elapsed unnoticed.

Certainly, his actions during the trip, like changing gears, slowdowns, and so on, were "just in time". Nevertheless, while all his movements occurred *on* time regarding the events, there was no representation *of* the time of the events. There was neither a perception of a "now", of a passed now, and of a now to come, neither the perception of a flowing of time. Events were not, in addition, put in a serial order, with a point of actuality beyond which there is a permanent anticipation of a net of possibilities for other events that will occur. In contradistinction, the attentive listener has a sharp perception of the time-order and of the time-flow of the sounds she listens to. This temporal organization of experience is not yet self-consciousness: rather, it is *the* consciousness of objects and events (with feelings, moods, proprioceptions, that are also events that join and are given together with the "external" events). Or, to put it differently, if we can talk of self-consciousness here, it has to do with the entrance of a zero-point of orientation for the organization of events which is centered on the "now" (this orientation-point is, really, totally subjective). Nevertheless, as a consciousness of the now, of time-order, and time-flow, this self-consciousness is fused with the events it seizes. It just sets the stage for the rise of a phenomenal world. This world is "for me", sure. But I am "out there", in the middle of it.

Second bizarre aspect: he learned (or enjoyed) nothing.

As a matter of fact, all his responses occurred "automatically" in the trip.

He was quiet capable, no doubt. Nevertheless, his mental states were simply running routines already established. No new ability, apprenticeship, enrichment of the driver's skill resulted from the trip. If there were in the journey some event not manageable with the routines established (a neon light that suddenly flashes in the night), the alert will sound in his mind, and this unusual event will be phenomenally seized as something happening now and showing something new that was in need of a deliberate response. On the contrary, the driver says to himself: "nothing new, everything as usual, the trip was quiet normal…" The temporal discrimination of events seems, thus, to have a close relationship with a center of decision able to rewrite routines or make entire new ones, as a learning process (my guess is that we also learn non-consciously, but in a rather slow, unremarked and cumulative way). This discrimination of events is not attention. Mental states truck-driver's-like are also attentional states. The other way around, events can appear in this decision-center without capturing mind's attention. The most striking situation is when we stare tediously at the passage of events in time, with nothing important to remark or to do. This center is rather something like a stage where events emerge and remain at disposal for inspection, direction and decision. It seems to be an interface, where perceptions, desires, beliefs, come together. It has a tremendous impact on the rapid adaptation of behavior. Nevertheless, almost all our behavioral connection with the surroundings has already begun in the deepest level of non-conscious mental life. This center is a flow of events that remain at disposal for active control. It is the flow of our conscious life. It is rather discontinuous (somnolence, sleeping, fainting, coma) and varying in intensity, but able to join the disparate parts into a delusive continuous flow. It is this operation that brings about the Cartesian illusion of a permanent, immaterial I, able to know, survey and control all his mental life.

Third bizarre aspect: in a sense, he was elsewhere.

Where? Precisely, where there was an experience developing with conscious mental states about objects (as defined above), and higher-order thoughts about him as entertaining those mental states. For instance, in his remembrance of past dinners with wife and kids, in his expectations about finally arriving at home, in the several thoughts that come to his mind while driving, as his longing for a beer at the next bar in the road and his decision to stop there. In a word, where there was an experience structured with the temporal organization of events at disposal of the interface where perceptions, remembrances or expectations mingle with desires, wills, beliefs, judgments, and so on, so that the "automaticity" of non-conscious mental states was substituted by a (very real) ability for pondering, inspecting, and making new decisions.

In a word, while having a point with his insistence that mental states may be non-conscious, Rosenthal does not seems to give an accurate account of

consciousness, so that Fisette's vindication of Brentano's intransitive and intrinsic consciousness is, for me, a simple question of respecting the facts of our mental life. Particularly, when Rosenthal says that, for a mental state, to be conscious is to be targeted by another mental state, the question about what consciousness is remains unanswered. A robot or a zombie could target is mental states by higher-order mental states. Nevertheless, all those states would be non-conscious. The property of being conscious for first-order mental states seems, for Rosenthal, to emerge by miracle with second-order mental states. It is hard to see why. We are still in need for an answer.

Looking into the other side, I ask Professor Fisette's why this consciousness must be from the outset construed as self-consciousness. It seems to me that it is rather a world-consciousness in the form of actuality (of course, with *qualia*). The fundamental is the position of a "now" (and a "here"), with its correlates, structuring the entire experience. They are actuality-makers and perspective-markers (not perspective-makers). Obviously, for a subject, to have before him an actual world structured with the subjective-dependent apprehension of a now and a here is tantamount to having a sense of himself as the center where a world-experience is displayed. But this sense *is not* yet to have a consciousness related to himself neither as an underlying subject, nor as a lateral consciousness of the mental states themselves. It is simply the experience of the world in such a form that it allows the later acknowledgment that this world, as it appears, is for him or from his point of view. Consequently, I think that the introduction of a perspective-point is legible on the things appearing and not on a kind of loop-like self-consciousness entrenched in the consciousness of things and events. It is the very conscious intentionality directed to things which is a perspective-laden relationship to those things. What is it like for a subject to see a red truck? I say: precisely seeing *this* red truck here and now. Where is the consciousness of seeing? It is in the thing actually seen. Then, self-consciousness really carves out a niche there! No. It would be superfluous for marking a perspective: the subjective point of view is already embedded in the way of seeing. But seeing red has a certain "feel", you say. Of course: seeing red is not seeing blue or seeing green. So, it is different for the subject to see red or to see blue, you insist. I agree: this is what sensing is all about – sensing is a discrimination ability that puts before me something as red, and something as blue, and so on; I do not need to sense my sensation in order to know that this is (or "feels", or "smells" like…) red. Husserl has a very deep insight about this. When he talked about the double intentionality of the flow of consciousness, he remarked that the retentional maintenance of the past appearances allows the apprehension of a temporal object right now, and, supplementary, the appearance of the subjective flow correlated with the temporal object. This "longitudinal intentionality" (*Längsintentionalität*) was, thus, the place where, for the very first time, we could grasp, over and above the red

object we saw, our own *seeing* of the red object. However, it is a second intentionality that we cannot confuse with the direct constitution of the temporal object and of its temporal phases: if we immerge into the intentional constitution of the temporal object, we see it as an objective unity of duration extending till the actual now, but we do not catch already our *seeing* of it.

What would I say?

The paper by Professor Fisette raises deep questions.

I wonder what image of human beings is conveyed by the Brentanian approach to mind. In a strong sense, Brentano is already a Cartesian, as he, defining mental states as conscious states, puts a clear dividing-line between the mental and the physical system that underlies it. Rosenthal's position, acknowledging the existence of non-conscious mental states, is to a certain extent beyond the Cartesian divide. Nevertheless, for both, the mental is taken as something with so clear a difference regarding the physical that one can always wonder how a physical system can "have" mental states. Physicalism, declaring that everything is physical or supervenes on it, says something that is certainly true, but, at the same time, poorly illuminating. If we question the philosopher of mind about what is a physical system, so that physical properties could be mental properties, he retorts to us: "well, ask the physicist". But, if we approach the physicist demanding what is physical reality in order for mental states to be (or supervene on) physical states, he will say for sure: "ask the psychologist". Nobody has an answer, and everybody thinks the other has.

In my opinion, the only productive starting point will be the whole living organism. The question would not be whether and how a physical system can have mental states, but how, in the global functioning of a living organism, we can trace a somewhat blurred line categorizing some functions as physical and other as mental. Intentionality is a productive, albeit tentative, response. If we define an intentional state as both a capability for triggering a response and/or mapping the surroundings (including the temporal and spatial position of the organism), we get a concept of intentionality that intersects what we usually conceive as the physical and the mental realms, so that the definition of an intentional state can function as a bridge or, better, as an overthrow of the traditional divide. A bio-chemical reaction in the cells is not an intentional state. But a net of chemical reactions that maps into the environment, so that, for instance, a response of the organism for escaping from fire follows, then it can be accounted for as an intentional state. In such a measure, we cannot be greedy when it comes to recognize an intentional state and, thus, a mind. A dog has perceptual states and beliefs (namely, that its owner is a reliable source of food); but a butterfly or a fly have very sophisticated systems for discriminating their environments and triggering adaptative responses to it.

Have they minds? If we feel inclined to say that they have, we are crossing the borders between a Cartesian-like conception of mind and an Aristotelian-like, where "souls" are not immaterial things with self-consciousness as an essential predicate, but instead patterns of high organization of bodily organisms.

Additionally, an organism does not live in the objective world (as described by natural science); an organism lives in a vital world, which is, we may say, a projection of its internal structure, a structure that results from a long and complex process of adaptation to its objective surroundings. The vital world of a human being is pretty different (and much more complex) from the vital world of a dog, a fly, or a bat. They all dwell in the same physical, objective world, certainly. But the kind of discriminations they can make, the type of behavior they can have, the kind of senses they use to scope around, the "pallete" of responses they are able to choose, all this is specific of each organism, and this specificity determines its vital world. So, a bat inhabits a unique vital world: the world of a bat. Only a bat could inhabit this vital world, and inhabiting it is all it is like to be a bat.

However, one may wonder, assuming the definition of an intentional state that I am sketching now is tantamount to acknowledging that there are mental states that will never be conscious states. This is precisely what I mean. Rosenthal rightly stresses that there are non-conscious mental states. Nevertheless, it seems that all non-conscious mental states can turn into conscious mental states by the intervention of a suitable HOT. I will risk the more radicalized idea that there are also mental states that cannot be conscious states at all. This is not a simple conjecture. Think of the well-known case of blind-sight. Then, think of such familiar situations as this one: when, suddenly, your arms move rapidly to protect your head from a ball that is coming to you, before you can even realize what is happening, you say that the response was made "by instinct"; the same with our truck driver, when he moves the steering wheel to avoid an obstacle, before becoming aware of what he is doing. In such situations, the perceptive system prompts a response that we can never be conscious of. We simply realize afterwards that we have act "without thinking", and we will never be in conditions to consciously recover the mental states we had, because they took place in an unconscious level. Considering that the top mental states, that can be conscious ones, float on a rather complex and intricate system of other mental states that will never be conscious, because they underpin the conscious ones and are covered by them, I propose to name them "pillar-mental-states", like the pillars of a bridge. We catch a glimpse of them when, such as in the examples I gave above, the underlying system goes its way faster than the related system of conscious mental states.

All in all, against Brentano and Fisette, I would say that consciousness is not the same as a mental state. Nevertheless, consciousness is not a superfluous or an elusive feature of our mental life. It is pretty real and terribly effective. It

controls behavior, allows comparison, ponderation, and fast learning processes. It displays a sense of subjectivity which is the most sophisticated mark of human life. Nevertheless, the building of a subjective point of view starts much sooner in the complexities of our organic life, so that, when we say "I", we are just expressing a point of view that grew out from a complex net of mental states that will never be accessible for reflection and introspection. They have prepared in advance a vital world where we, afterwards, consciously emerge as subjects of an actual experience. In a word, they are those "*petites perceptions*" or "*perceptions inaperçues*" Leibniz spoke about in one of the many "*aperçus géniaux*" of his brilliant mind.

I wish to express my gratitude to professor Fisette for his profound and suggestive paper, without which I will never be in conditions to get these (I guess) rather strange ideas.

ARGUMENTOS Revista de Filosofia

Denis Fisette*

Brentano's theory of consciousness revisited. Reply to my critics**

I am grateful to the respondents for the time and effort they have put into their comments on my paper on Brentano's theory of consciousness. Since many comments overlap, I will group them by theme and respond to the objections by clarifying certain aspects of my paper and by providing new elements in support of my reading of Brentano. First, I will justify the topic of my study in light of the various theories of mind that one can identify with the philosophy of Brentano. Second, I will summarize the main aspects of Rosenthal's HOT theory, which constitutes the background of my study. I will then discuss what since Chisholm has been called "Brentano's thesis," which many commentators defend in light of the two theses that I attribute to Brentano early in the target paper. In the fourth section, I will discuss several objections raised against my reconstruction of Brentano and his principle of the unity of consciousness. The main hypothesis that I developed in the last sections of the target article is that the later Brentano's introduction of the concept of mental agent aims at solving two main problems left pending in his 1874 *Psychology*. The first relates to the substrate of the modes of consciousness and of the complex mental state internally perceived. The second issue pertains to the status of the concomitant consciousness and to the second general thesis on consciousness: that all mental states are conscious. My hypothesis is that, to clarify the status of the *Mitbewußtsein* and to adequately answer the question: "What is it for a mental state to be conscious?," the later Brentano uses the concept of mental agent and conceives of consciousness as implicit and intransitive self-awareness.

The context. Brentano and contemporary philosophy of mind

This study focuses on the theory of consciousness that Franz Brentano develops in the second book of his *Psychology from an Empirical Standpoint* (1874) and also accounts for the later changes to his theory in his posthumously

* Université du Québec à Montréal. E-mail: denis.fisette@uqam.ca

** I wish to thank André Leclerc for his kind invitation to participate in this disputatio and Maxwell Ramstead for his careful linguistic revision of my paper.

published lectures and manuscripts. [1] My goal was to revisit Brentano's theory in light of divergent interpretations and understandings both in Brentano studies and in contemporary philosophy of mind. My starting point is the higher-order thoughts (HOT) theory of consciousness developed by the American philosopher David Rosenthal which, as I have shown in the target paper, has many affinities with Brentano's theory of consciousness. Rosenthal himself has repeatedly stressed the value of Brentano's contribution to the philosophy of mind,[2] and despite his disagreement with several aspects of Brentano's views on consciousness, he nevertheless considers that Brentano's theory shares much with his own.

G. Fréchette, P. Bernier, and A. Leclerc seem to have understood the purpose of my study differently when they complain that I did not consider several other options in contemporary philosophy of mind, options which, they believe, have more affinities with Brentano's view on consciousness than does Rosenthal's theory. To my knowledge, in addition to HOT theory and Kriegel's self-representational theory, which I also discussed in the target paper, there are at least three other theories of consciousness in contemporary philosophy that can be identified more or less explicitly with Brentano's. The first is the adverbial theory of consciousness, which dates back to the work of D. W. Smith (1986) and which has been recently advocated by A. Thomasson (2000, 2006) and A. Thomas (2003). In a nutshell, the adverbial theory maintains that awareness of one's mental state is expressed as an adverb in sentences such as "I present *consciously*," "I judge *consciously*," etc. We find the second option in the work of Tim Crane in philosophy of mind, particularly in his recent book *Aspects of Psychologism*, in which he associates "Brentano's thesis" (intentionality as the mark of the mental) to a (weak) form of intentionalism according to which "the nature of a conscious mental state is determined by its intentionality" (CRANE, 2013, p. 150; 2001, p. VII). Finally, Brentano's philosophy of mind is also associated to what U. Kriegel has recently called "The Phenomenal Intentionality Program" according to which "phenomenal intentionality is the intentionality a mental state exhibits purely in virtue of its phenomenal character". (KRIEGEL, 2013a; 2013b) This program is based on two theses that Kriegel also attributes to Brentano: intentionality is the mark of the mental, and all mentality is conscious (KRIEGEL, 2013b, p. 438).

All these options have the merit of showing the relevance of Brentano's theory of consciousness in light of current debates in philosophy of mind and they are based on solid knowledge of Brentano's writings. This is the case with Leclerc's intentionalism, on which he bases his commentary. There are many

[1] I use here the abbreviation "*Psychology*" to refer to the English translation of Brentano's *Psychologie vom empirischen Standpunkt* and "*Schriften* I" for the German edition of this book by Ontos (see the bibliography).
[2] Rosenthal repeatedly comments Brentano's theses on consciousness in his works. (see ROSENTHAL, 1991, 1993, 1997, 2003, 2005, 2009, 2011).

forms of intentionalism in philosophy of mind; the most radical calls for the reduction of consciousness to an intentional relation. But the form of intentionalism advocated by Leclerc is much weaker, since he recognizes the irreducibility of the two main characteristics that Brentano attributes to the mind (no consciousness without intentionality and vice versa), though Leclerc affirms the primacy of the intentional over the conscious. As we shall see, Brentano's theory differs from intentionalism or representationalism in that it assumes that an act as simple as hearing a sound involves several mental states that belong to multilayered classes standing in relations of dependence to one another. Hence the importance granted to the problem of the unity of consciousness in Brentano's descriptive psychology.

The fact that different theories can be identified with Brentano's philosophy clearly testifies that his theory of consciousness is open to interpretation. And indeed, my critics are aware that we also find quite different interpretations in Brentano studies, which range from "a neo-Brentanian theory of pre-reflective self-awareness" (BRANDL, 2013) to different versions of higher-order theories of consciousness.[3] I admit in the target paper that there are substantial differences between Brentano's and Rosenthal's theories, and I tried to show that Brentano's theory avoids several objections raised against HOT theory (GULICK, 2000). Nevertheless, beyond the differences and similarities that exist between these two theories of consciousness, Rosenthal's theory provides us with an appropriate theoretical framework for the reconstruction of the ins and outs of Brentano's theory of consciousness. I propose to clarify this point in the next three sections by summarizing, in the first, the main features of Rosenthal's HOT theory of consciousness, and by comparing it, in the second and third sections, to Brentano's theory.

HOT theory's main characteristics and the background

Rosenthal distinguishes two main traditions at the root of current trends in philosophy of mind, i.e., the Cartesian and the Aristotelian traditions. To each tradition corresponds a conception of the mind, and each can be characterized by using two key concepts of the philosophy of mind, namely intentionality and consciousness. The mind is defined in the Cartesian tradition by consciousness, while Aristotelianism favours intentionality (1990, p. 735). Rosenthal (1986, p. 332, 335-336) claims that the conception of mind advocated in these two traditions determines their respective conception of consciousness.

3 Fréchette even casts doubt on the number and relevance of references to the theories of higher order of consciousness in Brentano studies. Yet this rapprochement is at the heart of numerous well-known studies that I mentioned in the target paper, the most important in this regard being those from Caston (2002), Zahavi (2004, 2006) and more recently Gennaro (2012).

Rosenthal further distinguishes two notions of consciousness: state consciousness and what he calls "creature consciousness," i.e., the consciousness of an organism or a subject. Attributed to a state, the predicate "is conscious" simply refers to the property of a mental state of being conscious. The notion of creature consciousness designates a property of the agent herself, one that varies as a function of whether, e.g., she is awake or in a coma. Cartesianism seeks to answer the question: "What it is for a creature to be conscious?" while the main question raised by Aristotelianism is: "What it is for a mental state to be conscious?" A third important distinction is that between two different uses of the attribute "is conscious": an intransitive one, which does not require a direct object (such as being conscious or unconscious, to be anxious, cheerful or excited, etc.) and a transitive one, which requires a direct object (e. g. being aware of the noise, etc.). Transitive consciousness is another name for intentional consciousness and refers to the relation that an agent bears to something other than herself. In its intransitive use, "is conscious" refers to a monadic predicate that stands for a non-relational property (ROSENTHAL, 1990, p. 737). To this distinction between transitive and intransitive uses of "is conscious" there corresponds a distinction between two types of properties that are attributable to a mental state: intrinsic and extrinsic properties. This distinction finds its linguistic expression in the distinction just made between transitive and intransitive uses of this predicate. When construed as a monadic predicate, it refers to an intrinsic property; when used as a relation, it designates an extrinsic property. In Rosenthal own terms:

> A property is intrinsic if something's having it does not consist, even in part, in that thing's bearing some relation to something else. If being conscious is at least partly relational, a mental state could be conscious only if the relevant relation held between the state and some other thing. (ROSENTHAL, 1990, p. 736).

In an Aristotelian conception of the psychical, where consciousness is not essential to mental states, consciousness is considered an extrinsic property.

With the help of these four distinctions, one can provide an explicit definition of both traditional conceptions of consciousness. For Cartesianism, consciousness is an intrinsic, intransitive and non-relational property of mind, while for Aristotelianism, on the contrary, consciousness is defined as an extrinsic and transitive property of a mental state (ROSENTHAL, 1990, p. 737). Moreover, in conceiving of consciousness as an intransitive and intrinsic property of the person or creature, Cartesianism presupposes that the subject is aware of all his thoughts or all the contents of his mental states.[4] That is why,

[4] In support of his diagnosis, Rosenthal quotes a passage from Descartes' *Meditations* (fourth replies), in which Descartes claims that "aucune pensée ne peut exister en nous sans que nous en soyons conscients au moment même qu'elle existe en nous" (Descartes, Quatrièmes Réponses, *Œuvres de Descartes*, édition

according to Rosenthal, Cartesianism deprives us of the ability to provide a non-circular explanation of consciousness by conflating two distinct questions: that of "a state's being conscious" and that of "one's being conscious of that state" (ROSENTHAL, 1997, p. 735; 2009, p. 4; 1986, p. 337). Methodologically, Rosenthal's theory proceeds in reverse order to Cartesianism, in that it considers that our answering the question: "What it is for a mental state to be conscious?" is a prerequisite to answering the question of creature consciousness. What is specific to Rosenthal's theory is the idea that the consciousness of a state mainly depends on the intentional relation between a HOT and the initial state it is targeting. According to Rosenthal:

> We are conscious of something, on this model, when we have a thought about it. So a mental state will be conscious if it is accompanied by a thought about that state. […] The core of the theory then, is that a mental state is a conscious state when, and only when, it is accompanied by a suitable HOT. (ROSENTHAL, 1990, p. 741).

A HOT is a purely intentional state which, contrary to a HOP (higher order perception) in D. Armstrong's model, has no qualitative property. (Rosenthal, 2005, p. 105) Its two main proprieties are its propositional content and its assertoric mode or attitude. The propositional content that accompanies a state of pain, for example, takes the following form: "I now have (or feel) a pain in my stomach". This thought must have an assertoric mode because to make the target state conscious, one must posit the existence of that state and, more precisely, posit that one is in this state (ROSENTHAL, 1991: 31; 2009, p. 2). A sensory state that is not accompanied by a HOT cannot be considered a pain because, as we said, this sensory quality does not pre-exist the thought that we have. Finally, as I have shown in the target paper, higher order thoughts are numerically distinct from the lower order, generally unconscious states towards which they are directed (ROSENTHAL, 1997, p. 742).

Now, even if one admits that the thought accompanying the initial state makes that initial state conscious, we can still ask what it is for a mental state to be conscious at all. For, to attribute consciousness to a mental state, it is necessary to presuppose that one is oneself in the target state because, as P. Alvez pointed out, the mental state, taken for itself, cannot be said to be aware of anything. In other words, conscious states are those mental states that are one's own. I can only be aware of my own stomach pain, and not of someone else's. Being transitively conscious of the target state is a relation that a creature bears to that state. The higher order thought must therefore be about the fact that one is oneself in that mental state (ROSENTHAL, 2002, p. 658; 1997,

de Ch. Adam et P. Tannery, v. VII, Paris, J. Vrin, 1964-1965, p. 246; on Brentano's debt to Descartes' philosophy, see D. Fisette, 2015).

p. 738, 740-741; 1986, p. 344). The content of a HOT, then, could tentatively be formulated in the following way: "I am now in a state of fear, anxiety, etc.; I am now feeling pain". Hence the principle of transitivity, which Rosenthal formulates as follows: "the view that a state's being conscious consists in one's being conscious of that state" (ROSENTHAL, 2009, p. 4; 2005, p. 4).

In her commentary, Perez points out that Rosenthal's theory is but one possible version of higher order (HO) theory of consciousness in general, and wonders if Brentano's theory has not more in common with another version of HO developed recently by P. Carruthers. The latter is actually a version that differs from Rosenthal's HOT theory in that the target state is not an actual but a potential state which is conscious "by virtue of being *disposed* to give rise to a higher-order thought". According to Carruthers' dispositionalist HOT theory of consciousness,

> a conscious mental event *M*, of mine, is one that is disposed to cause an activated belief (generally a non-conscious one) that I have *M*, and to cause it non-inferentially (CARRUTHERS, 2007, p. 13).

Although Brentano admits of unconscious dispositions in his *Psychology*, they play no role in his theory of consciousness. In this regard, Brentano agrees with Rosenthal that the initial state, that is, the primary object (e.g., my hearing of a song), has to be the *actual* object of inner perception.

Brentano' s theory of consciousness and HOT

Brentano's commentators are divided as to whether the conception of mind that he defends in his *Psychology* makes him a Cartesian or an Aristotelian.[5] According to the received view (mainstream at least since R. Chisholm), Brentano's key concept is intentionality, a concept that he had the merit of reintroducing into the vocabulary of contemporary philosophy. Hence, "Brentano's thesis" (so-called) which, as I have already noted, is the common starting point of Leclerc, Fréchette, Bernier, and intentionalist theories of consciousness. And indeed, several passages in Brentano's work during his Vienna period seem to support this interpretation. For example, early in the second chapter of his *Psychology*, Brentano denounces the ambiguity of the term "consciousness" and uses it to designate the property of a mental state's being about an intentional object (*SCHRIFTEN* I, p. 119).

> I prefer to use it [the term consciousness] as synonymous with "mental phenomenon," or "mental act". For, […] the term "consciousness," since

[5] Notice that in some of his works, Rosenthal also associates Brentano with the Cartesian camp (ROSENTHAL, 1990, p. 746-747; 1991, p. 30; 1993, p. 211-212; 2004, p. 30 f.; 2009, p. 4).

> it refers to an object which consciousness is conscious of, seems to be appropriate to characterize mental phenomena precisely in terms of its distinguishing characteristic, i.e., the property of the intentional in-existence of an object, for which we lack a word in common usage. (*PSYCHOLOGY*, p. 78-79).

We find a similar remark in *The Origin of our Knowledge of Right and Wrong*:

> The common feature of everything psychical consists in what has been called by a very unfortunate and ambiguous term, consciousness; i.e., in a subject-attitude; in what has been termed an *intentional* relation to something which, though perhaps not real, is none the less an inner object of perception (*innerlich gegenständlich gegeben*). (BRENTANO, 1902, p. 12).

However, a closer examination of the chapters of his *Psychology* devoted to consciousness reveals that consciousness and intentionality, although coextensive, stand for two distinct properties of mental states. These two properties correspond to the two theses that I attribute to Brentano at the very beginning of my paper: every psychical phenomenon is consciousness of something (*Bewußtsein*) and every mental phenomenon is conscious (*bewußt*).[6] I argued that Brentano's theory of primary and secondary objects aims at articulating these two main theses.

Rosenthal clearly saw that, in emphasizing state consciousness (in the second thesis) over creature consciousness and in conceiving of consciousness as an (intrinsic) property of mental states, Brentano occupies a position in between the Cartesians and the Aristotelians. Rosenthal (2009, p. 2) maintains that the originality of Brentano's theory over the tradition of Descartes and Locke rests on the idea that every mental state is conscious (thesis II)[7] and on the explanation he provides "both of what it is for states to be conscious and of why, as he held, all mental states are conscious". (ROSENTHAL, 2009, p. 2) This interpretation complements his theory of primary and secondary objects, in which mental phenomena are understood as secondary "objects" that are in principle the only ones that can be internally perceived in the first edition of Brentano's *Psychology*. The study of this thesis is the main subject of the second book of Brentano's *Psychology*, and at the outset, he opposes this thesis to the hypothesis of unconscious mental states, which is one of the main postulates of Rosenthal's theory of higher order thoughts.

[6] At the very beginning of his lecture on descriptive psychology, Anton Marty explicitly refers to these two theses in order to characterize Brentano's conception of the mental in his descriptive psychology. (MARTY, 2011, p. 9).

[7] Rosenthal says later in this article that "it was rare until Brentano's time to describe mental states as conscious at all. Even though Descartes and Locke were plainly writing about the property we describe as a state's being conscious, they did not say that our mental states are all conscious, but rather that we are conscious of all our mental states". (ROSENTHAL, 2009, p. 4).

Fréchette and Bernier have misgivings with regard to the distinction between these two theses. Fréchette wrongly accuses me of advocating the thesis that "Brentano's account of consciousness makes consciousness a relational (or transitive) feature of the mind," while Bernier does not see why consciousness in Brentano cannot be both transitive, as required by the first thesis, and intransitive, in the sense that the predicate "is aware" would be a monadic predicate designating an intrinsic and non-relational property of mental states. In fact, Bernier claims that "there is simply no contradiction in claiming that the predicate is conscious can be used both transitively to say of a subject that she is conscious of something and intransitively to talk of a mental state by virtue of which the subject is conscious of something". Of course, there is no contradiction if the predicate is used transitively in relation to a *mental state* and intransitively in relation to a *creature*. However, in the passage of my paper to which Bernier refers in that context, I say only that these two uses of the predicate in question cannot be applied simultaneously to one and the same isolated *state*. Leclerc also questions this dual use of the predicate "is aware" and wonders in what sense consciousness of the secondary object can be described as intransitive and intrinsic because, according to Leclerc, "having an object" is part of the *definiens* of what we call "intentionality".

These objections raise an important question about the interpretation of Brentano's second thesis on consciousness, namely that of the status of the concomitant awareness, about which he repeatedly says in his *Psychology* that it accompanies each and every mental state. The difficulty arises from the interpretation of a mental state's being conscious in terms of its being an object of consciousness. This difficulty stands out clearly in the famous passage of Brentano's *Psychology*, in which he wrote:

> We can say that the sound is the *primary object* of the *act* of hearing, and that the act of hearing itself is the *secondary object*. Temporally they both occur at the same time, but in the nature of the case, the sound is prior. […] The act of hearing appears to be directed toward sound in the most proper sense of the term, and because of this it seems to apprehend itself incidentally (*nebenbei*) and as something additional (*als Zugabe*). (*PSYCHOLOGY*, p. 198). [8]

The terms *nebenbei* (incidentally) and especially *Zugabe* (additional) suggest that the concomitant awareness that accompanies the presentation of the sound is something extrinsic to hearing and merely constitutes an additive, like cream or sugar added to coffee: and in this sense, the concomitant

[8] Compare this passage with the following excerpt from his lectures on descriptive psychology: "Every consciousness, upon whatever object it is primarily directed, is concomitantly directed upon itself (*geht nebenher auf sich selbst*). In the presenting of the colour hence simultaneously a presenting of this presenting. Aristotle already [emphasizes] that the psychical phenomenon contains the consciousness of itself". (BRENTANO, 1995, p. 25).

awareness would be imposed from without on a mental state, just as in Rosenthal's theory, the content of the higher order thought makes the target state conscious. This hypothesis cannot be rejected out of hand when one takes into account certain aspects of Brentano's psychology that are presupposed in his theory of consciousness. I have in mind the rapprochement, which Brentano makes in his *Psychology* (p. 22, 70), between concomitant consciousness and inner perception, the latter of which is defined there as a judgment, i.e., as an attitude (*Stellungsnahme*) and as cognition. Several commentators of Brentano, especially Leclerc and M. Textor (2013b), maintain that the concomitant consciousness in Brentano is a judgment, more specifically an immediately evident cognition of the primary object. Textor correctly argues that, although judgment in Brentano is assertoric and has a function similar to that of a HOT, namely the function to posit the existence of the primary object, it remains that internal perception is not a categorical judgment, but rather an immediate and evident existential judgment. (*SCHRIFTEN* I, p. 161-163) This is what Textor calls the Dual Relation Thesis (DRT):

> every mental phenomenon M is primarily directed upon an object other than M and secondarily (concomitantly) upon M itself in a way that yields knowledge of M. (2013b, p. 446).

DRT emphasizes the epistemic function of internal perception and amounts to reducing consciousness to a kind of cognition. DRT seems to presuppose that the judicative mode underlying the epistemic function of internal perception is the only mode by which *consciousness* relates to its objects. But there are reasons to believe that Brentano distinguishes the epistemic functions of consciousness from the non-epistemic ones.

First, in a footnote to the title of the first chapter, "Inner Consciousness," Brentano explicitly distinguishes inner consciousness from internal perception: "Just as we call the perception of a mental activity which is actually pres ent in us "inner perception," we here call the consciousness which is directed upon it "inner consciousness". (PSYCHOLOGY, 1995, p. 68).

Second, at the very beginning of the third chapter of the second book of his *Psychology*, after having established that every mental act is accompanied by a concomitant consciousness, i.e., that in hearing a sound, for example, the presentation of that sound is always accompanied by a consciousness of itself, Brentano says that mental phenomena are the modes or ways by which consciousness enters into relation with its objects. (*Psychology*, p. 107) This implies that judgment is only one of the three possible modes by which one becomes aware of an object: representational, judicative and emotional. The mode of relation to the object that includes only a presentation is the poorest and merely consists in the fact that the object is present to consciousness. The other two modes suppose the active stance of consciousness with regard to its

objects. They are characterized by the opposition, in the intentional relation of judgment and emotions to their respective objects, between ascent or affirmation and negation in the case of judgments, and love and hate in the case of emotions. Internal consciousness therefore has an extension broader than internal perception, understood as judgment (in its epistemic function), since it applies equally to all classes of mental states, including to that of presentations.

Third, according to one of the principles at the basis of Brentano's classification of acts, the class of presentations is not only the simplest of acts, but is also ontologically independent of the class of judgments. This means that the presentation of the presentation of the sound or the consciousness that accompanies this presentation is not necessarily a cognition. In this regard, remember that Brentano clearly distinguishes the hierarchical relation between the three classes of acts from that between primary and secondary objects. For, in the first case, the relation of foundation between the first class and the other two leads to a one-sided (*einseitig*) dependence of judgments and emotions on the class of presentations, which is in principle autonomous with regard to the two remaining classes. However, between the consciousness of the primary object and the consciousness of the secondary object, there is a bilateral (*gegenseitig*) relation of dependence, in the sense that both *relata* are mutually dependent. This bilateral dependence is a presupposition of the two general theses on consciousness. (BRENTANO 1954, p. 226-227).

Finally, in his Vienna lectures on descriptive psychology, Brentano provides further information about his analyses on consciousness in his *Psychology* and introduces some distinctions that seem to argue in favour of the distinction between the epistemic and the non-epistemic senses of consciousness. I am thinking especially of the distinction between implicit awareness (or consciousness broadly understood) and explicit awareness (consciousness in the narrow sense), the latter of which is closely associated with the central concept of noticing (*Bemerken*) in these lectures. Brentano first applies this distinction to the external perception of a primary object and argues that, not only can one implicitly see or hear something that one does not explicitly see, but that to be explicitly aware of experiencing something, one must be implicitly aware of it (BRENTANO 1995, p. 36-37). Explicit consciousness or secondary consciousness is called in these lectures a noticing (*Bemerken*) (BRENTANO 1995, p. 36-37), which roughly corresponds to the epistemic mode of consciousness in Brentano's *Psychology*. Brentano also applies this distinction to self-awareness and we shall see that explicit awareness seems to presuppose a form of reflection in Brentano's later writings.

Nevertheless, I recognize that the judicative mode of consciousness is the one that Brentano emphasizes in his *Psychology* and in several other writings because of its epistemic function. But for the reasons I just mentioned, this

epistemic function is distinct from its psychological function. That is why, I believe, DRT leaves completely open the status of these modes of consciousness or the concomitant awareness that is supposed to accompany the primary object. This problem stands out clearly in Brentano's response to the question raised by thesis II, i.e., "What it is for a mental state to be conscious?" Brentano's response in his *Psychology* is simply that my hearing the sound is object of consciousness, which is not only circular but also vulnerable to the objection of the "consciousness of" trap raised against HOT theory. Brentano's response has also been challenged by Leclerc in his commentary, where he questions the nature of the mode of consciousness at work in relation to the secondary object. If, as he suggested, this relation were intentional, then the secondary object could be reduced to the initial presentation of the primary object and the secondary consciousness would stand in an intentional relation to the presentation of the primary object. However, as Brentano clearly saw, this option would lead to an infinite regress:

> As I have already emphasized in my *Psychology from an Empirical Standpoint,* however, for the secondary object of mental activity one does not have to think of any particular one of these references, as for example the reference to the primary object. It is easy to see that this would lead to an infinite regress, for there would have to be a third reference, which would have the secondary reference as object, a fourth, which would have the additional third one as object, and so on. (*PSYCHOLOGY*, p. 215).

The later Brentano has made substantial changes to his conception of secondary objects and we shall see that these changes go hand in hand with the abandonment of the concept of concomitant consciousness in favour of that of self-awareness.

My reconstruction of Brentano's theory and the principle of the unity of consciousness

Now, in spite of numerous parallels that can be drawn between the theory of consciousness developed by Brentano in his *Psychology* and Rosenthal's HOT theory, there are important differences as well, which I have stressed in the target paper. The two main differences pertain to two postulates of HOT theory of consciousness, namely that of unconscious mental states and the "distinctness assumption," i.e., that the target state and HOT are two numerically separate states. Brentano discusses these two assumptions in connection with his second thesis on consciousness, which is exposed to objections of infinite regress well-known since Aristotle. For, when one denies that the presentation that accompanies the hearing of the sound is unconscious, as most higher order theories of consciousness hold, it seems that we are then forced to admit

an infinite number of mental phenomena. I have argued in the target paper that Brentano's answer to this objection is that the presentation of the sound and the presentation of the presentation of the sound are one and the same state that has two objects, a primary and a secondary object. There are no unconscious presentations in the field of our experience, nor can there be (*PSYCHOLOGY*, p. 81), and the objection of infinite regression is not an argument against his theory, because the series of acts ends with the second term. (*PSYCHOLOGY*, p. 100).

The key to Brentano's solution to the problem of regression (*Psychology*, p. 98) lies ultimately in the idea of a special connection (*eigentümliche Verwebung*) between the primary and secondary objects and it raises once again the question of the nature of this relation. I tried to show that, for Brentano, the consciousness of the primary object and the consciousness of the secondary object are metaphysical parts, or what Brentano called in his *Psychology* "divisives," of a single unitary phenomenon, and they are part of one and the same reality. Hence the principle of the unity of consciousness, which Brentano already evokes in the first chapter of Book II of his *Psychology* in response to the question why the many mental phenomena that are involved in the simplest acts appear to consciousness not as an aggregate or bundle of dispersed elements, but as a unitary reality. It is in this context that Brentano uses his theory of wholes and parts, and conceives of mental phenomena as "parts of one single phenomenon (*Teilphänomene*) in which they are contained, as one single and unified thing". (*PSYCHOLOGY*, p. 74) This principle is not intended to eliminate the complexity of mental acts in favour of simplicity, but aims rather at warranting that what is internally perceived is a unitary whole. The principle of the unity of consciousness also asserts that all mental states involved in this complex act are also perceived simultaneously. (*PSYCHOLOGY*, p. 171; *Schriften* I, p. 182-183; 1995, p. 125-126).

All respondents seem to agree with my reconstruction of this aspect of Brentano's theory, and de Carvalho provides further useful information on other aspects of Brentano's psychology that are involved in his theory of consciousness. Most objections pertain to the relation that I have established between the principle of the unity of consciousness and the mental substance in Brentano's revised theory, which I will later discuss. In his contribution to this *disputatio*, B. Leclercq examines the nature of the dependence relations involved in Brentano's ontological solution to the problem of the unity of consciousness. He claims that the relations that Brentano establishes between the distinctional parts of a complex unitary mental act require a richer ontology than the one developed by Brentano in his *Psychology*, and that this ontology has been developed by his student Husserl and formalized recently by G. Null (2007). Bruno Leclercq emphasizes the distinction between two classes of relations of dependence: the first class is "relative dependence," which obtains

between two interdependent parts of a whole, where one of the parts is "more fundamental" than the other; the second class of dependence is said to be "weaker" than the first because it only supposes that relations of dependence obtain independently of the founded-founding relations that hold among the parts. Leclercq argues that the class of relative dependence could help establish the priority of consciousness over intentionality. Yet, even if one agrees with Leclercq's proposal, it is difficult to see how this distinction could contribute to the problem of the unity of *consciousness*. Be that as it may, all I needed in order to underpin Brentano's ontological solution to this problem during his Vienna period were the bilateral distinctional parts in the proper sense, i.e., that primary and secondary objects are mutually inseparable. But Leclercq certainly knows that the later Brentano developed a sophisticated theory of relations, to which he made several changes during his career. (CHRUDZIMSKI & SMITH, 2004). As we shall see, these changes are important for Brentano's revised theory consciousness.

Maria Gonzalez and Mariana Broens discuss Brentano's principle of the unity of consciousness through what I have called in the target paper the problem of complexity, i.e., the problem of unifying within inner consciousness the entire complex of elements involved in the constitution of our mental life. The original but complicated solution they propose to this problem involves a combination of information theory (Dretske), ecological psychology, and Complex Systems Theory (CST). One of the properties of CST that seems relevant to account for the unity of consciousness is self-organization, which is understood here "as a process through which new forms of organization emerge solely from the dynamic interaction amongst elements". They argue that "both primary and secondary objects of consciousness can be understood as having the same informational nature" and that the unity of consciousness can be accounted for "from the informational perspective enriched by Complex Systems Theory". And this presupposes, in turn, that Brentano's theory of primary and secondary objects can be accounted for in terms of information (and meaning), which they conceive of along ecological lines, i.e., as ecological "invariant features of the world" including affordances, niches, etc. In this perspective, "meaningful information emerges in consciousness as a result of the agent's adaptive interaction with the environment".

Gonzalez and Broens' proposal raises the issue of whether Brentano's theory of primary and secondary objects is compatible with this ecological worldview. They are aware that their proposal is based on a conception of mind which is known for its anti-representationalism and the question arises as to how it fits in not only with Brentano's psychology but with his metaphysics as well. Pedro Alvez has raised a similar question in his criticism of Brentano's principle of the unity of consciousness, but unlike Gonzalez and Broens, he argues that one must choose between the two conflicting options. For he

conceives of the "soul" as "patterns of high organization of bodily organisms" and maintains that this view represents an antidote to Brentano's Cartesian dualism and an alternative to his reprensentationalist conception of the mind. Although I am sympathetic to Gonzalez-Broens' overall perspective in their paper, I must also acknowledge with Alvez that their proposal raises insuperable difficulties in light of the objections that Husserl and several of Brentano's followers have raised against his descriptive psychology. Alvez's diagnosis is based on Brentano's internalism and mentalism, and he argues that this form of representationalism is simply inconsistent with the role assigned to the environment and the body in later phenomenology and in ecological psychology. The burden of proof lies therefore with Gonzalez and Broans. As for the solution they propose to the problem of complexity, it all depends on the type of relations that are involved in the emergence of these forms of organization. This is not the place to discuss that difficult issue. Nevertheless, let me remind the reader that several of Brentano's students were strongly interested by this issue in their work on Gestalt psychology, which in turn is one of the sources of Gibson's ecological psychology.

Self consciousness and the mentally active agent

Most objections raised by Perez and especially by Bernier and Fréchette relate to the last two sections of the target paper, in which I assess the implications of the changes in Brentano's philosophy for his theory of consciousness. The main hypothesis that I developed in these two sections is that his taking into account the concept of psychical agent aimed to solve two major problems left open in his 1874 *Psychology*. The first issue pertains to the question of the substrate of the modes of consciousness and of the complex psychical act as internally perceived. The hypothesis that there is such a substrate has raised numerous objections, which I will discuss in the last section. The second problem, which I discussed in the previous section, is related to the status of the concomitant consciousness and to the second general thesis on consciousness (that every mental state is conscious). For, as I have repeatedly stressed in my contribution to this volume, to the question of what it is for a mental state to be conscious, Brentano responded in his *Psychology* by saying that it is the secondary object of a concomitant consciousness that accompanies the initial state, understood as its primary object. The predicate "is conscious" is therefore not an intrinsic property of mental states, as Bernier and Fréchette claimed in their commentaries, because for Brentano the consciousness of mental state depends upon the *Mitbewußtsein* that takes them as objects. My hypothesis is that to clarify the status of the accompanying awareness and to adequately answer the question: "What it is for a mental state to be conscious?," the later Brentano resorts to the concept

of mentally active agent and conceives of consciousness as implicit and intransitive self-awareness.

Fréchette strongly disagrees with this hypothesis and proposes his own intentionalist-unilevelist interpretation of Brentano, which can be summarized by the thesis that "Brentano shares with Rosenthal the assumption that state consciousness is a primitive fact, and that it explains creature consciousness". We shall see that Rosenthal and Brentano claim that a correct definition of consciousness involves both state and creature consciousness. In any case, Fréchette does not admit that the introduction of a psychical agent into the later Brentano's philosophy changes anything about his theory of consciousness: "After all," Fréchette adds, "instead of talking about 'consciousness', and preferring 'mental agent' or 'mental activity', the basis of Brentano's account remains, at bottom, unchanged in his later view". In other words, the only difference that he sees between consciousness and mental agent is a mere *façon de parler*. Frechette's main argument is based on his own exegesis of Brentano and his strategy consists in casting doubt on the authenticity of those writings of Brentano (namely *Religion und Philosophie*) that I quoted to support my hypothesis. This is clear from his interpretation of the well-known 1911 passage, to which many Brentano scholars usually refer in order to explain the important modifications to which Brentano's views on consciousness were subject after 1874. (TEXTOR, 2013b, p. 479-480). In this passage, Brentano maintains that the secondary object is no longer a mental state being about itself (*in parergo*) as a secondary object, as he held in his *Psychology*, but rather the mentally active subject that includes the primary and secondary object:

> As I have already emphasized in my *Psychology from an Empirical Standpoint*, however, for the secondary object of mental activity one does not have to think of any particular one of these references, as for example the reference to the primary object. It is easy to see that this would lead to an infinite regress, for there would have to be a third reference, which would have the secondary reference as object, a fourth, which would have the additional third one as object, and so on. The secondary object is not a reference but a mental activity, or, more strictly speaking, the mentally active subject, in which the secondary reference is included along with the primary one. (*PSYCHOLOGY*, p. 215; *SCHRIFTEN* I, p. 395).

According to Fréchette, whereas in 1874, Brentano claimed that "every conscious act contains a primary and a secondary object," he maintained in 1911 that "the mentally active subject includes both the primary reference (my seeing red) and the secondary reference (my being conscious of seeing red)". Fréchette does not seem to realize that "my being conscious of seeing red," which presupposes that it is the *creature* that is conscious of the primary object, is quite different from *state* consciousness and also from the idea that the predicate "is conscious" is intrinsic to the state of seeing red. For, if we take Fréchette's formulation at face value, Brentano would have shifted from

Aristotelianism (state-consciousness) to Cartesianism (creature consciousness). And what is worse, he accuses me of advocating the idea that consciousness in the work of the later Brentano requires that he "introduces level (3) to address these issues," as is necessary in HOT theory. While it is true to say that I claim that Brentano recognizes in 1911 that a response to the question: "What makes a mental state conscious?" must necessarily take into account the creature or the mental agent, I also claim that Brentano avoids the drawbacks associated with creature consciousness by conceiving of consciousness in terms of self-consciousness. Brentano can thus preserve his second thesis on consciousness (that any mental state is conscious) while providing an explanation that, as can be shown in the excerpt from the appendix, is different from the 1874 explanation, where it was understood as a mere (secondary) "object" of a *Mitbewußtsein*. In any case, one can hardly deny that for the later Brentano one of the main conditions imposed to thesis II (that all mental states are conscious) is that the mental agent be conscious of it. And this requirement can be considered a clarification of the obscure notion of concomitant consciousness that was supposed to accompany the initial state (e.g., the hearing of a sound) in Brentano's first edition of his *Psychology*. Fréchette therefore minimizes the extent of the modifications, which are mentioned in the passage of the appendix and which can be observed in several of Brentano's later writings. I shall now comment on these briefly.

I am thinking of Brentano's writings gathered under the title *The Theory of Categories* (1981), which he wrote during the last ten years of his life and which seem to corroborate the passages of *Religion und Philosophie* that I used in the target paper. We find further illuminating remarks about the connection between the concepts of psychical agent, which he construes as a mental substance, and self-awareness defined as "a cognition (*Kenntnis*) which pertains to that which has the cognition" (BRENTANO, 1981, p. 116). But this definition has to be nuanced by means of the distinction between implicit and explicit consciousness that Brentano introduced in his lecture on descriptive psychology, which I used above in my discussion of the Dual Relation Thesis.[9] Brentano also associates the distinction between implicit and explicit consciousness to that between broad and narrow consciousness or to that between blind and distinct consciousness, which is closely related to the central concept of noticing (*Bemerken*) in these lectures. Implicit and indistinct consciousness characterizes primary consciousness, while explicit and distinct consciousness is understood as secondary consciousness. Brentano first applies this distinction to the external perception of a primary object, arguing that one can see or hear (implicitly) something that one does not (explicitly)

[9] See K. Mulligan (2004) for a discussion of these distinctions in Brentano's lectures on psychognosy.

perceive. In *The Theory of Categories*, Brentano uses the case of the hearing of a chord in order to exemplify this distinction:

> If one hears a chord and distinguishes the notes which are contained in it, then one has a distinct awareness of the fact that he hears it. But if one does not distinguish the particular notes, then one has only an indistinct awareness of them. In such a case, he does hear them together and he is aware of the whole which is this hearing and to which the hearing of each of the particular notes belongs; but he does not hear the whole in such a way that he distinguishes each of its parts. Particular hearings of particular notes are contained in the whole and he does not distinguish them. (BRENTANO, 1981, p. 117).

But this case, like that of the lark that I use in the target paper, only concerns the (primary) consciousness of the primary object. In another passage of his *Theory of Categories*, Brentano also uses this distinction in his analysis of self-awareness by taking the example of pain:

> Self-awareness, too, is sometimes distinct and sometimes indistinct. If a person feels a pain, then he is aware of himself as one that feels the pain. But perhaps he does not distinguish the substance, which here feels pain, from the accident by means of which the substance appears to him. It may well be that animals have only an indistinct self-awareness. But in the case of man, the substance which thinks in him [*die in ihm denkt*], and experiences, judges, loves and hates, can be brought to awareness as a result of the frequent change of its accidents; the indistinct awareness is then replaced by a distinct awareness of the subject. One then grasps this substance as that which permanently underlies this change and which gives unity to its manifold character [*als das, was bleibend ihrem Wechsel und einheitlich ihrer Mannigfaltigkeit unterliegt*]. (BRENTANO, 1981, p. 117).

As in the case of the hearing of a chord, one has to presuppose that the subject is aware of the fact that he hears it, and in the case of pain, that she is aware of being in that state. Although Brentano does not admit of unconscious mental states, he assumes here that explicit self-consciousness presupposes implicit self-awareness, and so confirms the thesis of his lectures on descriptive psychology, to wit, that one cannot be explicitly aware of being in this state (pain) unless one is implicitly aware of it (BRENTANO 1981, p. 34). This implicit self-awareness is not reflexive; it does not require, as Brentano says (1981, p. 123), the participation of the will. It is therefore pre-reflective, i.e., an awareness that one has before explicitly reflecting on one's experience, and it is intrinsic.

Brentano and the principle of transitivity

Bernier's detailed commentary provides useful insights in Brentano's theory of consciousness from the perspective of contemporary philosophy of mind. He takes for granted my interpretation of the later Brentano's theory of

consciousness in terms of pre-reflective and intransitive self-consciousness, and claims that Brentano's revised theory can be understood along the lines of a one-level representationalist theory of consciousness. To quote Bernier:

> The mental agent would stand in a representational, and hence intentional, relation both to the primary object and to the secondary object, namely herself, or herself mentally acting. According to such an interpretation, however, the mental agent could no longer be a "unified real being" since it would only be "intentionally existent", like the primary object.

Bernier's point of departure rests on the mereological interpretation of the relation between primary and secondary objects that I formulated in the target paper, and he rightly points out that this formulation primarily holds for Brentano's early theory of consciousness but not necessarily for the revised theory. He then proposes several formulations of Brentano's revised theory which take into account the function of the mentally active subject. According to Bernier, the following statement captures the gist of the later Brentano's theory as I presented it in the target paper:

> 4*. For any state M of a subject S, M is a mental state of S iff M is conscious, where M is conscious iff M is an act of S such that by M-ing S represents a primary object O, and S is non-intentionally, directly aware of herself and of her M-ing.

Bernier further argues that even if Brentano's theory so understood has the virtue of accounting for the phenomenal subjective character of conscious states, it still carries "an ontological burden" (the mental substance) which is not necessary to a Brentanian or neo-Brentanian theory of consciousness. He then proposes several variants of an ontologically neural theory of consciousness and concludes, that while these versions depart significantly from Brentano's theory as formulated in 4*, "these views can still be called Brentanian, or neo-Brentanian, in the important sense that they all correspond to what has been called 'one-state views' in the literature".

Bernier might be right to say that Brentano's philosophy conveys numerous metaphysical presuppositions from the naturalistic standpoint adopted by most contemporary philosophers of mind, including Bernier himself, but as Leclerc rightly pointed out, this is not an argument to discard Brentano's theory of mind as a whole. For Brentano's psychological concern in referring to a mental agent in the revised theory, first and foremost, was to account for the conscious character of a mental state, and this concern is distinct from Brentano's ontological considerations on the mental substance. I shall return to the ontological issue in the next section. As for Bernier's statement 4*, it seems to presuppose that Brentano advocates a Cartesian conception of the mind and commits himself to creature consciousness. Again,

Brentano's revised theory does not involve a shift from state to creature consciousness and he neither conflates nor confuses these two concepts. Nevertheless, this raises the question how the reference to a subject makes it possible to answer the initial question pertaining to what makes a mental state conscious. In this regard, Bernier's first version of statement 4 seems more relevant for that purpose because the thesis that figures in the antecedent of the conditional rests on conscious mental states whereas this is not the case in 4*, which is oriented towards the condition for being a mental state simpliciter. Second, statement 4* does not account for the fact that M is a complex state which includes the relation to the primary and secondary objects, as Brentano makes clear in the 1911 excerpt that I quoted above. Third, to account for the idea that the mentally active agent is conscious of being in that complex state, the formulation that Bernier proposes of the consequent of the conditional, a formulation in terms of the subject being "non-intentionally, directly aware of herself and of her M-ing," has to be modified. For, the subject does not experience herself in the same way that she is conscious of the primary object; rather, she is aware of being in that complex state, which includes both the primary and the secondary object. Nevertheless, I agree with Bernier that the mereological definition that I proposed in statement 3, which was only meant to account for the ontological structure of the complex state, has to be completed in order to account for the subject's awareness of being in that complex state.

Now, Bernier and Perez have rightly pointed out that the main issue is whether or not Brentano's revised theory of consciousness implements Rosenthal's transitivity principle (TP). As Bernier rightly remarked, my position on that issue in the target paper is "a bit unstable," mainly because of my formulation of this principle using the terms "conscious of " and my transitive use of the predicate "is conscious," neither of which are to be confused with TP. We find a clear formulation of TP in the introduction to Rosenthal's book *Consciousness and Mind*, where he repeatedly insists on the importance of this principle for HO theories of consciousness in general. This formulation of TP is broad enough to accommodate several versions of higher order theories of consciousness. TP: Mental states are conscious only if one is in some way conscious of it (2005, p. 4; 2009, p. 4).

This principle states the conditions for a mental state to be conscious and it involves the idea that the predicate "is conscious" is attributable to a mental state only if the subject is somehow conscious of that state. Rosenthal claims that this principle is common to all HO theories, which mainly differ in the way they implement this principle. We saw how HOT theory implements this principle by accounting for the for-me-ness of lower-order states. It consists in the thesis that having a HOT that one is in some state consists in being conscious of oneself as being in that state. (2005, p. 6) Rosenthal also believes that higher order thoughts are unconscious in that we usually do not notice

that we are in those states. That is why he resorts to third-order thoughts, by which the subject *explicitly* becomes aware of the content of the state she is in. (ROSENTHAL, 1990, p. 742) Since Brentano denies the very idea of unconscious mental states, the question remains whether his revised theory implements TP.

Perez and Bernier adopt contradictory positions on this issue. Bernier claims that if statement 4 discussed above is correct, then Brentano's revised theory implements TP. But even if we agree with Bernier that the general condition that a state must satisfy to be conscious is that the subject "be non-intentionally, directly aware of herself and of her mental act," this doesn't explain how Brentano's theory is supposed to implement TP. Perez, on the other hand, raises doubts as to whether the very idea of consciousness in Brentano is compatible with TP, insofar as Brentano's revised theory does not satisfy the main conditions generally imposed on HO theories, namely the distinctness assumption and the postulate of unconscious mental states. In particular, she asks whether Brentano's notion of implicit consciousness is vulnerable to the main arguments raised against HOT theory that she discusses in her commentary. She seems to believe that the only way out is through the adoption of a pre-reflective first order theory of phenomenal consciousness. As I said above, I don't think that first-orderness is an issue in the interpretation of Brentano's theory of consciousness because unlike a one level or a one state view, the content of an elementary experience such as the vision of a colour is complex and multi-layered in Brentano's account.

One of the main issues raised by the objections directed against most HO theories is whether (self-) consciousness pre-exists psychical acts such as the presentation of a sound. This question underlies Alvez's and Perez's discussion of D. Armstrong's distracted driver case. Alvez argues against Brentano that the truck driver case, far from being marginal, is paradigmatic of the way we behave most of the time. He further claims that unconscious mental states are the most important part of our mental lives and "that there are also mental states that cannot be conscious states at all". This last claim is difficult to justify from Brentano's empirical standpoint. For, even if Brentano would grant that a mental state could forever remain implicit, he would not accept that a mental state could not be potentially raised to consciousness. In the target paper, I used a similar case from Brentano's lecture on descriptive psychology to illustrate the distinction between implicit and explicit consciousness, i.e., the two ways that a mental state can be an object of consciousness. I have argued that this distinction, like the one between marginal and focal consciousness, allows us to account for the truck driver case without resorting to the unconscious. For one can experience something like a lark in the visual field or the notes of a chord in the hearing of a musical piece without being explicitly and distinctly conscious of it. But unlike Alvez's hypothesis of unconscious mental states or contents, Brentano would say that

for a state to be a mental state, it has to be somehow experienced or be a datum of the agent's experience.[10]

Be that as it may, the question whether Brentano's theory of consciousness implements Rosenthal's TP presupposes that this principle constitutes an adequate criterion to identify a HO theory and to discriminate the latter from non-HO theories. For to take the creature into account in this formulation still does not explain the role of self-awareness in the agent's experiencing the primary object.[11] It only shows that:

- A mental state is conscious iff the mentally active subject is somehow conscious of that state.

I think we need more to account for self-consciousness. For this formulation does not seem to take into account the fact that mental states are the agent's own, i.e., in Brentano's terms, that the initial presentation is not merely a state but a state that the subject is in. Moreover, we have to account for Brentano's important remark in the 1911 passage, that what the secondary consciousness stands in relation to is not an object as such, but rather the mental agent, in which both the intentional object and the state are included. I take it that Brentano means that the hearing of a sound is the state the agent is in and that a state is conscious only if she is conscious of being in that state. In short, an adequate response to the question: "What makes a state conscious?" could be formulated along the following lines:

- A mental state is conscious iff one is aware of oneself as being in that (intentional) state.

[10] André Leclerc discusses similar cases in his commentary, but he adopts a position diametrically opposed to those advocated by Alvez and Perez, who argue that the solution to the majority of these problems requires the adoption of phenomenal consciousness. According to Leclerc, the theoretical framework that Brentano established in the first edition of his *Psychology* provides all the necessary elements to address these problems, provided that Brentano be interpreted from an intentionalist perspective. Among the problems typical of representationalist theories of mind, Leclerc mentioned the cases of pain and of several mental states, such as anxiety or moods, to which many philosophers refuse to attribute intentional properties, because they believe that they are objectless states which are not about anything. In the case of pain, I think Brentano would agree with Leclerc that they are intentional states, as confirmed by Brentano in the extensive discussion he devoted to this question in his *Psychology* (p. 62-69) and in his polemic with Stumpf on the status of pleasure and pain. (FISETTE, 2013b). Brentano believes that cases like pain fall under the class of emotions, which, like any intentional state, intentionally in-exist. Leclerc takes the example of the phantom limb as a case of non-conceptual and sensorial experience and argues that, like states with conceptual content, a pain can be about something which does not exist. But Leclerc's argument presupposes that an itch felt by somebody in a non-existing part of one's body is nevertheless the intentional object of his pain. This presupposition is questionable because the localisation of pain in one's body part does not necessarily account for the aboutness or directedness of a state of pain. For one can be in a state of pain without being able to localise the source. What is then the object of pain?

[11] Caston also believes that the solution to the problem of consciousness presupposes creature consciousness, i.e., "*our* perceiving that we perceive"; "It is not, therefore, mental states like perceptions that are aware, strictly speaking, but rather the animals themselves who have these mental states". (CASTON, 2002, p. 769).

This formulation seems to be consistent with Brentano's conception of self-consciousness in his final writings, namely in the passages of his *Theory of Categories* that I quoted above, where he formulates his conception of self-consciousness by using the distinction between implicit and explicit consciousness. But unlike de Carvalho, who claims, after Brandl (2013), that the idea of self-consciousness is already in the first edition of Brentano's *Psychology*, in the next section, I will explain why the concept of a self does not play any role in the 1874 theory of consciousness.

The later Brentano on substance and accident

Let us now discuss the second part of the overall hypothesis that I stated above, according to which the introduction of the notion of mental agent in Brentano's revised theory of consciousness is associated not only to changes in his conception of substance and accidents, but also to his solution to the problem of the substrate of the modes of consciousness, which he left pending in his *Psychology*. Fréchette attaches great importance to this issue, if I consider his numerous objections to this aspect of my target paper, which says very little about the ontological status of the self and the mental substance. In fact, this issue was not essential to my overall argument on Brentano's later theory of consciousness, and that is why this aspect of the target paper was very sketchy. However, the question remains whether, after the reist turn, the immaterialist conception of the soul, which Brentano contrasts with Aristotle's alleged semi-materialism, has anything to do with the modifications brought to his theory of consciousness. I will try to meet Fréchette's objections and provide further textual information on the most important points.

First, Fréchette claims that "nothing in the text used by Fisette [the excerpt from *Religion und Philosophie* on Aristotle's semi-materialism] is actually referable to Brentano's 'late position', since it is composed of and/or inspired by numerous texts by Brentano (and Marty) belonging to different unidentified periods". It is true that Brentano's writings published in *Religion und Philosophie* are undated, and like many of Brentano's later writings published by O. Kraus, A. Kastil and F. Mayer-Hillebrand, that piece is not entirely reliable given the editorial policy adopted by Marty's students in their editions of Brentano's writings. (FISETTE, 2013a). Nevertheless, Fréchette should know that I could have used several other manuscripts where Brentano criticizes Aristotle's semi-materialism, namely *Vom Dasein Gottes* (1980, p. 424 f.) and above all Brentano's manuscripts published in *The Theory of Categories* that I quote above and that undoubtedly belong to Brentano's final period (1907-1917). Fréchette is right to say that Brentano's metaphysical position on substance changed several times over the years (CHRUDZIMSKI, 2004), but I take for granted that Brentano's conversion to an immaterialist conception of substance

occurred during the later period of his philosophical activity. Moreover, these manuscripts have been authenticated and used by several Brentano scholars, namely by Antonelli in the introduction to his recent edition of Brentano's *Psychology* (SCHRIFTEN, p. LXXX) and by S. Krantz (1988) in reference to Brentano's later criticism of Aristotle.

Fréchette further claims that "Brentano never doubted that there is a substrate to our conscious mental states. This substrate is called the soul". Here again, Brentano would disagree, as is clear from the position he defended in his *Psychology* and even as early as 1869, in his paper "Auguste Comte und die positive Philosophie". Brentano criticizes Aristotle for conveying metaphysical presuppositions in a number of his doctrines, notably in that of substance and accidents.[12] Brentano raises the same objection at the very beginning of his *Psychology*, when he compares the Aristotelian conception of psychology as a science of the soul to the one defining it as the science of mental phenomena. Brentano criticizes Aristotle for conceiving of the object of psychology, that is, the soul, as a substance, and of psychical phenomena as its accidents or its essential properties. Brentano argues that, from an empirical point of view, this is nothing but a metaphysical postulate, i.e., a fiction, which, because it is not (and cannot be) an object of experience or an object accessible to internal consciousness, consequently cannot constitute the object of psychology. Hence the alternative conception, based on a "psychology without a soul," i.e., a psychology free of metaphysical presuppositions. (FISETTE, 2014b) It is probably for the same reason that Brentano, in the conclusion to his analysis of the unity of consciousness, deliberately left open the question of the substrate and individuality of mental states, arguing that the unity of consciousness and the unity of the conscious self are two distinct things:

> Finally, the unity of consciousness does not imply that the mental phenomena which we ordinarily refer to as our past mental activities, were parts of the same real thing that encompasses our present mental phenomena. […] It remains an open question, then, for the moment, whether the continued existence of the self is the persistence of one and the same unitary reality or simply a succession of different realities linked together in such a way that, so to speak, each subsequent reality takes the place of the reality which preceded it. (*PSYCHOLOGY*, p. 129-130).

In this regard, Brentano's lecture on descriptive psychology marks a return to Aristotle and to a psychology understood as an ontology of the soul.

[12] F. Brentano (1968, p. 132): "Aristotle who, despite being a theist, is not a theological thinker (in the erroneous sense), despite depending on metaphysical conceptions in a number of his doctrines, such as those of potency and act, of substance of accident, etc.—this, even his greatest admirer cannot deny. He is nevertheless already a positive researcher by his character. Up until him, there is an order similar to the one Comte determines in a general manner. Consequently, we should have expected a refinement and more perfect development of the positive spirit".

(BRENTANO, 1995, p. 155) The "*letzteinheitliche Subjekt*" in *Religion und Philosophie* (p. 227), the self, is then considered an individual substance whose moments or properties are mental states. In his *Theory of Categories*, Brentano maintains that the mental substance is not a mere a priori postulate but an object of experience insofar as "each of us is conscious of himself as being a determinate individual and as being the one individual substance that underlies all our psychical activities" (BRENTANO, 1981, p. 121).

Fréchette once again misunderstands my position when he says that "the immortality, or even the existence of the soul, was a condition for the unity of consciousness". Fréchette claims instead that Brentano's point is "that the unity of consciousness is what *makes* a being (a creature) conscious". I must say that I cannot understand how this principle of the unity "of consciousness" could possibly make a state or a human creature "conscious". In the target paper, I argued that this principle was intended to solve the problem of complexity, i.e., the problem of why the various phenomena that are involved in the simplest acts appear to consciousness not as an aggregate of scattered elements, but as a unitary reality. The other condition that is associated with this principle is the simultaneity condition, according to which one must be aware that this multiplicity of elements belongs to one and the same reality. In this sense, the simultaneity condition is to the consciousness of the unitary phenomenon, what the ontological condition of membership to one and the same reality is to the object internally perceived. However, this principle does not address the question of what makes a mental state or a creature conscious. Bernier also errs when he says: "The mental agent, however, is supposed to be what plays the role of unifying the diverse parts of the mental states". As I said, Brentano's adoption of the concept of self-awareness as well as his taking into account the experience of the subject call into question neither the validity of his theory of primary and secondary objects nor the central function of the principle of the unity of consciousness in his overall conception of the mind. As the real substrate of all modes of consciousness, the mental substance is the seat of the unity of consciousness, but it is not its unifying principle.

This is confirmed by Brentano's remarks on Aristotle's semi-materialism. In a passage of *The Theory of Categories*, Brentano first explains why he characterizes Aristotle's position on substance as semi-materialistic:

> I have said that our self appears to us as a mental substance. I now add that it appears to us as a pure mental substance. It does not appear, say, as a substance which is mental with respect to one part and which is corporeal, and thus extended in three dimensions, with respect to another part. I emphasize this expressly, for the contrary has been asserted by important philosophers - for example, by Aristotle in ancient times and by many present-day thinkers who have been influenced by his opinion. (BRENTANO, 1981, p. 121-122).

The purely mental substance entails that the subject underlying the mental states is an immaterial substance, insofar as it is neither part of the body nor of the brain, and is free of any spatial properties. (BRENTANO, 1954, p. 226) Aristotle, on the other hand, can be considered a semi-materialist (or semi-immaterialist), insofar as he conceived of the soul "as a composition of corporeal and un-corporeal parts" and attributed "to the different parts of our sensory perceptions and desires different parts of the corporeal subject". (BRENTANO, 1954, p. 224).

The next question raised by Fréchette pertains to the connection between the unity of consciousness and Brentano's *letzteinheitliche Subject*. Fréchette sees no link because he believes that Brentano's view on substance "doesn't play any role in the phenomenological fact of the unity of consciousness". It is true that this principle is distinct from Brentano's metaphysical views on substance, but there is nevertheless a connection. Indeed, one of Brentano's arguments against Aristotle's semi-materialism rests on the fact that Aristotle's conception of substance infringes the principle of the unity of consciousness. Brentano is categorical on this point, as shown in this passage from *Religion und Philosophie* (BRENTANO, 1954, p. 227, 224):

> [Aristotle] doubly infringes the secured fact of the unity of consciousness. First by conceiving the soul as a composition of corporeal and uncorporeal parts. Second, by attributing to the different parts of our sensory perceptions and desires different parts of the corporeal subject.

I cannot examine in details Brentano's argument. Nonetheless, this passage makes it clear that this principle presupposes a conception of the soul as an immaterial substance. I see another connection between the mentally active subject and the principle of the unity of consciousness, more specifically, in the requirement of simultaneity, according to which the phenomena involved in the activity of the subject should appear to consciousness as a unitary reality. For, in Brentano's revised theory of consciousness, the simultaneous consciousness (*gleichzeitige Gesamtbewußtsein*) is the whole whose parts are the ultimate unitary subject's (letzteinheitliche Subjekt) own mental states (BRENTANO, 1954, p. 225, 227). It follows that the unitary consciousness of the whole is a self-consciousness and the whole is the self who is himself distinctly or indistinctly apperceived through his own parts, as we have seen above, and as confirmed in another passage of Brentano's *Theory of Categories*:

> And if he thinks or senses indistinctly, then the self is comprised in a larger complex which is at least apperceived as a whole, even if not in respect to its relevant particular parts. In such a case one has a confused self-awareness with no distinction of the relevant particular psychical activities. (1981, p. 123).

Fréchette might be right to say that the justification of the principle of the unity of consciousness is based on internal perception and not on the mental substance. However, the content of what is internally perceived in the later Brentano always involves the self.

References

ARMSTRONG, D. "What is Consciousness?", In: BLOCK N., FLANAGAN O., and Güzeldere G. (Eds.). *The Nature of Consciousness:* Philosophical Debates, Cambridge: MIT Press, 1997, p. 721-728.

BRANDL, J. (2013) "What is Pre-Reflective Self-Awareness? *Brentano's Theory of Inner Consciousness Revisited*". In: FISETTE D., and FRÉCHETTE G. (Eds.). *Themes from Brentano*, Amsterdam: Rodopi, p. 41-66.

BRENTANO, F. (*Schriften* I). *Sämtliche veröffentlichte Schriften*, v. 1, *Schriften zur Psychologie, Psychologie vom Empirischen Standpunke* / Von der Klassifikation der Psychischen Phänomene, T. Binder and A. Chrudzimski (Eds.), Frankfurt a. M.: Ontos, 2011.

_____. *Psychology from an Empirical Standpoint*. Transl. by A.C. Rancurello, D. B. Terrell, and L. McAlister. London: Routledge, 1973.

_____. (2003). *Vom Ursprung sittlicher Erkenntnis*. Leipzig: Dunker & Humblot, 1889; *The Origin of the Knowledge of Right and Wrong*, Trans. by R. Chisholm and E. Schneewind, London: Routledge, 1969.

_____. (1995). *Descriptive Psychology*. Transl. and ed. by B. Müller, London: Routledge; trans. of: *Deskriptive Psychologie*, R. Chisholm and W. Baumgartner (Eds.), Hamburg: Meiner, 1982.

_____. *The Theory of Categories*. Transl. R. Chisholm and N. Guterman, The Hague: Nijhoff, 1981.

_____. Vom Dasein Gottes. A. Kastil (Ed.), Hamburg: Meiner, 1980.

_____. « *Auguste Comte und die positive Philosophie* », *Chilianeum*. *Blätter für katholische Philosophie, Kunst und Leben*, v. 2, 1869, p. 15-37; reprinted In: O. KRAUS (Ed.). *Die vier Phasen der Philosophie und ihr augenblicklicher Stand, nebst Abhandlungen über Plotinus, Thomas von Aquin, Kant, Schopenhauer und Auguste Comte*, Hamburg: Felix Meiner Verlag, p. 97-133, 1968.

_____. *Religion und Philosophie*. F. Mayer Hillebrand (Ed.), Bern: Francke, 1954.

_____. *Wahrheit und Evidenz*. O. Kraus (Ed.), Leipzig: Meiner, 1930.

_____. *The origin of our Knowledge of Right ans Wrong.* Trans. C. Hague, Wesrminster. Archibald Constable & Cie, 1902.

CARRUTHERS, P. Higher-order Theories of Consciousness. *Stanford Encyclopedia of Philosophy,* 2007. (online).

CASTON, V. "Aristotle on Consciousness", *Mind*, v. 111, 2002, p. 751-815.

CHALMERS, D. "Facing Up to the Problem of Consciousness". *Journal of Consciousness Studies*, v. 2, n. 3, 1995. p. 200-219.

CHRUDZIMSKI, A. *Die Ontologie Franz Brentanos*, Dordrecht: Kluwer, 2004.

_____. Smith, "Brentano's Ontology: From Conceptualism to Reism". In: JACQUETTE, D. (Ed.). *The Cambridge Companion to Brentano*, Cambridge: Cambridge University Press, p. 197–219, 2004.

CRANE, T. *Aspects of Psychologism*, Harvard, Harvard University Press, 2013.

_____. *Elements of Mind*. *An Introduction to the Philosophy of Mind*. Oxford: Oxford University Press, 2001.

FISETTE, D. (2015). « Le « cartésianisme » de Franz Brentano et le problème de la conscience », In: S. Roux (Dir.). Le corps et l'esprit: problèmes cartésiens, problèmes contemporains. Paris: Éditions des archives contemporaines, p. 163-208.

_____. (2014a). Duas teses de Franz Brentano sobre a consciência. *Phainomenon. Revista de Fenomenologia*, v. 22-23, p. 9-30.

_____. (2014b). "Franz Brentano et le positivisme d'Auguste Comte", no spécial de la revue *Cahiers philosophiques de Strasbourg*, v. 35, n. 1, p. 85-128.

_____. (2013a). "Introduction to Section V: Expositions and Discussions". In: D. Fisette and G. Fréchette (Eds.). *Themes from Brentano*, Amsterdam: Rodopi, p. 359-268.

_____. (2013b). "Mixed feelings". In: FISETTE, D., and FRÉCHETTE, G. (Eds.). *Themes from Brentano*. Amsterdam: Rodopi, p. 231-306.

GENNARO, R. *The Consciousness Paradox:* Consciousness, Concepts, and Higher-Order Thoughts, Cambridge: MIT Press, 2012.

GULICK, R. van. "Inward and Upward. Reflection, Introspection, and Self-Awareness". *Philosophical Topics*, v. 28, p. 275-305, 2000.

KASTIL, A. *Die Philosophie Franz Brentanos. Eine Einführung in seine Lehre*, Bern: Francke, 1951.

KRANTZ, S. F. "Brentano's argument against Aristotle for the immateriality of the soul". *Brentano Studien*, v. 1, p. 63-74, 1988.

KRIEGEL, U. (Ed.). *Phenomenal Intentionality*. Oxford: Oxford University Press, 2013a.

_____. "Phenomenal intentionality: past and present, introductory". *Phenomenology and the Cognitive Science*, v. 12, p. 437-444, 2013b.

MARTY, A. *Descriptive Psychologie*, M. Antonelli & J. C. Marek (Eds.), Würzburg: Königshausen & Neumann, 2011.

MULLIGAN, K. "Brentano on the Mind". In: JACQUETTE, D. (Ed.). *The Cambridge Companion to Brentano*. Cambridge: Cambridge University Press, 2004, p. 66-97.

NULL, G. "The ontology of intentionality", two parts. *Husserl Studies*, v. 23, p. 33-69; 119-159, 2007.

ROSENTHAL, D. "Concepts and Definitions of Consciousness." *Methodology and History of Psychology*, v. 4, n. 3, p. 55-75, 2009.

_____. *Consciousness and Mind*. Oxford: Oxford University Press, 2005.

_____. "Varieties of Higher Order Theory". In: R. Gennaro (Ed.). *Higher-Order Theories* of Consciousness. Amsterdam: John Benjamin Press, 2004, p. 17–44.

_____. "Thinking that One Thinks". In: DAVIES, M. and HUMPHREYS, G. W. (Eds.). *Consciousness*. Oxford: Blackwell, p. 197-223, 1993.

_____. "The Independence of Consciousness and Sensory Quality". *Philosophical Issues*, v. 1, p. 15-36, 1991.

_____. "A theory of consciousness". BLOCK, N. *et al*. (Eds.). *The Nature of Consciousness. Philosophical Debates*. Cambridge, MIT Press, 1990, p. 729-753.

_____. "Two Concepts of Consciousness". *Philosophical Studies*, v. 49, n. 3, p. 329-359, 1986.

_____. and H. LAU. "Empirical support for Higher-Order Theories of Conscious Awareness". *Trends in Cognitive Sciences*, v. 15, n. 8, p. 365-373, 2011.

SMITH, D. W. "The structure of (self-)consciousness". *Topoi*, v. 5, n. 2, p. 149-156, 1986.

TEXTOR, M. "Neither a Bundle nor a Simple: Brentano on the Unity of Consciousness". In: D. FISETTE and G. FRÉCHETTE (Eds.). *Themes from Brentano*. Amsterdam: Rodopi, 2013a, p. 67-86.

_____. "Brentano on the dual relation of the mental". Phenomenology and the cognitive Science, v. 12, p. 465-483, 2013b.

_____."Brentano (and some Neo-Brentanians) on Inner Consciousness". *Dialectica*, v. 60, p. 411-431, 2006.

THOMAS. "An adverbial theory of consciousness". *Phenomenology and the Cognitive Sciences,* v. 2, p. 161-185, 2003.

THOMASSON, A. (2000). "After Brentano: a one-level Theory of Consciousness". *European Journal of Philosophy*, v. 8, p. 190-209.

_____. "*Self-Awareness and Self-Knowledge*". *Psyche,* v. 12, n. 2, p. 1-15, 2006.

ZAHAVI, D. "Two takes on a one-level account of consciousness". *Psyche*, v. 12, n. 2, p. 1-9, 2006.

ARGUMENTOS
Revista de Filosofia

VARIA

Peter Simons*

Wittgenstein and surprise in mathematics

"In great mathematics there is a very high degree of unexpectedness".
G. H. Hardy

"[T]he mathematician is not a discoverer: he is an inventor."
Wittgenstein

ABSTRACT

One of the psychologically strongest motivations for mathematical platonism is the existence of surprises in mathematics. Time and again results have turned up which went contrary to the expectations of even the best qualified. Wittgenstein was always an anti-platonist, so for him there could be no surprising discoveries about mathematical objects as there can be about animals in the Amazon basin or chemicals on Titan. Given the later Wittgenstein's algorithmic conception of mathematics, it might appear that for him the only legitimate notion of surprise in mathematics must be merely psychological. In this paper I examine whether a less subjective conception is compatible with his position in the philosophy of mathematics.

Keywords: Wittgenstein; mathematics; platonism; surprise.

RESUMO

Uma das mais fortes motivações, em termos psicológicos, para o platonismo matemático é a existência de surpresas em matemáticas. Com frequência, resultados apareceram que foram contrários a expectativas de até mesmo os mais qualificados. Wittgenstein sempre foi um anti-platonista, então para ele não pode existir descobertas surpreendentes sobre objetos matemáticos como pode haver sobre animais na bacia amazônica ou sobre produtos químicos em Titan. Partindo-se da concepção algorítmica do Wittgenstein tardio, deve parecer para que a única noção legítima de surpresa na matemática deve ser uma meramente psicológica. Neste artigo, eu examino se uma concepção menos subjetiva pode ser compatível com a sua posição em filosofia da matemática.

Palavras-chave: Wittgenstein; matemática; platonismo; surpresa.

* Trinity College Dublin. Email: psimons@tcd.ie

Compulsion and surprise

Two phenomena conspire to convince people that the physical world exists independently of them. One is its recalcitrance, or insusceptibility to control. It resists and constrains our actions. Much as we might wish to do so, we cannot lift heavy boulders, walk through walls, jump rivers, breathe under water, or fly (unaided) over mountains. The other feature, which is connected to the first, is the world's propensity to surprise us. The sights and sound, pressures and pains of the world force themselves upon us in perception whether we want them to or not, and are often unexpected and surprising. An unusual bird appears in the garden, a stranger calls at the door and reveals he is a long-lost cousin, the post brings an invitation out of the blue, the car won't start (surprises may be unpleasant as well as pleasant). These two phenomena, recalcitrance and surprise, form a large part of the platonist's case for the existence of an independent mathematical reality. The recalcitrance of mathematical reality indeed appears to be stronger than that of the physical: the necessity with which mathematical results follow from assumptions is stricter than the physical necessity by which a wall resists attempts to walk through it. This has rarely been put more eloquently than by the Polish logician Jan Łukasiewicz. Speaking in particular of mathematical logic, he wrote

> whenever I work on even the least significant […] problem, […] I always have the impression that I am facing a powerful, most coherent and most resistant structure. I sense that structure as if it were a concrete, tangible object, made of the hardest metal, a hundred times stronger than steel and concrete. I cannot change anything in it; I do not create anything of my own will, but by strenuous work I discover in it ever new details and arrive at unshakeable and eternal truths (LUKASIEWICZ, 1970, 249).

One of the most difficult tasks for an anti-platonist, such as Wittgenstein, is to explain this sense of confronting a recalcitrant independently existing reality. And to the end of accounting for the appearance of mathematical necessity and giving it a non-platonist explanation, Wittgenstein devoted much attention.

The phenomenon of surprise in mathematics is also frequently cited as evidence for the independence of mathematical existence. Though it is less widely discussed than the notion of compulsion in mathematics, its persuasive power is if anything greater than that of necessity. Mathematical necessity is pervasive, and mathematicians and commentators on the subject are so used to it that it takes an apparent exception to it to grab their attention. Such an exception is afforded by such results as the independence of the continuum hypothesis, or earlier, the consistency of non-Euclidean geometries. Such exceptions are instances of mathematical surprise, and there are others in the history of mathematics. The most famous is the discovery of incommensurable

numbers, a surprise which may even have cost the discoverer his life.[1] Others, more directly relevant to Wittgenstein's intellectual *milieu*, were the paradoxes of set theory, and the incompleteness of arithmetic discovered by Gödel. Because surprises are salient, they provide dramatic phenomenological evidence for the mind-independence of the mathematical. It is therefore incumbent on an anti-platonist like Wittgenstein to find an alternative explanation for the phenomenon of surprise in mathematics, one which rejects the idea that mathematical objects are out there waiting to be discovered, like so many unvisited planets.

Surprise and the surprising

Whether or not the most penetrating, certainly the most charming philosophical account of surprise that I know is provided by a section of an early work by Adam Smith, a history of astronomy published only posthumously in 1795. Distinguishing surprise from wonder at the novel and admiration of the great, Smith remarks that it is the unexpectedness of what is discovered that constitutes its peculiar feature:

> When an object of any kind, which has been for some time expected and foreseen, presents itself, whatever be the emotion which it is by nature fitted to excite, the mind must have been prepared for it, and must even in some measure have conceived it before-hand; because the idea of the object having been so long present to it, must have before-hand excited some degree of the same emotion which the object itself would excite: the change, therefore, which its presence produces comes thus to be less considerable, and the emotion or passion which it excites glides gradually and easily into the heart, without violence, pain, or difficulty.
> But the contrary of all this happens when the object is unexpected; the passion is then poured in all at once upon the heart, which is thrown, if it is a strong passion, into the most violent and convulsive emotions, such as sometimes cause immediate death; sometimes, by the suddenness of the extacy, so entirely disjoint the whole frame of the imagination, that it never after returns to its former tone and composure, but falls either into a frenzy or habitual lunacy; and such as almost always occasion a momentary loss of reason, or of that attention to other things which our situation or our duty requires. (SMITH, 1967, p. 32).

Smith then goes on to show with graphic examples how dramatic the effects of surprise can be. As this shows, surprise is a psychological reaction to the unexpected, which in intensity may range from mild to overwhelming, indeed sometimes so overwhelming as to prompt disbelief in the supposed datum.

[1] The story is that Hippasus of Metapontum inadvisedly made or announced his discovery while at sea, and was thereupon thrown overboard by scandalized fanatical Pythagoreans.

Surprise is to be distinguished from being surpris*ing*. Something which is usually not itself mental is surprising if it surprises the first people who come upon it or discover it, or which typically surprises those who come across it for the first time in their own experience even after it has become known. It was a surprising discovery that life teems around deep-sea vents: no one, not even experts, had expected there to be life, let alone an abundance of life, in the pitch black of the ocean deep. Something which typically surprises those who experience it for the first time is the size of St. Peter's basilica in Rome. No matter how much they have seen pictures of it, the scale when one is present in person is greater than one would expect. Both of these examples depend on the prior and independent existence of the object in question. So if there is anything in mathematics which is surprising in either sense, if the analogy between the physically surprising and the mathematically surprising holds, it is evidence for the mind-independent existence of mathematical objects.

Wittgenstein on surprise

At no point in his philosophical career was Wittgenstein prepared to endorse the platonist conception of mathematics. The *Tractatus* is brief about mathematics, but since according to it there are no genuine mathematical propositions, the question of what they are about does not arise. At any rate, "there can never be surprises in logic" (6.1251). Wittgenstein does not go on to say whether there can be surprises in mathematics, but given his tendency to treat logic and mathematics on a par in the *Tractatus*, we must assume he would think there cannot be genuine surprises. In any case, in the sense of surprise being the reaction to something unexpected by those who first come upon it, he was wrong about logic, at least second-order logic. The incompleteness results of Gödel were genuinely surprising at the time, even at first to Gödel, and the hints in the *Tractatus* that there could be a mechanical method for deciding which propositions were logically valid were soon shown by Church to be unfounded even for first-order logic. Wittgenstein had genuine misgivings about Gödel's result, and while his sniping at Gödel's proof is not one of his more impressive efforts, his doubts were shared at the time by more technically versed logicians such as Zermelo and Leśniewski.

Wittgenstein dealt with the notion of the surprising in mathematics in a series of remarks, left out of the first edition of *Remarks on the Foundations of Mathematics* (without explanation) and inserted in the revised edition of 1978 (again without explanation or apology). Their juxtaposition (as Appendix II of Part I) with some of his remarks on Gödel (Appendix III thereof) add weight to the idea that the surprising in mathematics was perceived as a challenge to his anti-platonism.

Wittgenstein distinguishes two roles that surprise can play in mathematics:

> "The surprising may play two completely different parts in mathematics".
> "One may see the value of a mathematical train of thought in its bringing to light something that surprises us:—because it is of great interest, of great importance, to see how such and such a kind of representation of it makes a situation surprising, or astonishing, even paradoxical.
> "But different from this is a conception, dominant at the present day, which values the surprising, the astonishing, because it shews the depths to which mathematical investigation penetrates;—as we might measure the value of a telescope by its shewing us things that we'd have had no *inkling* of without this instrument. The mathematician says as it were: "Do you see, this is surely important, this you would never have known without me." As if, by means of these considerations, as by means of a kind of higher experiment, astonishing, nay *the most* astonishing facts were brought to light."
> "But", protests Wittgenstein immediately, "the mathematician is not a discoverer: he is an inventor." (111)

The first role of surprise is a legitimate one, but it is presentational only: by leading up to a result in a certain way it is highlighted as surprising. The unstated implication is that, were the result presented differently, it would not be surprising. No example is given to illustrate how this can occurs, but here is a possible candidate for the sort of thing Wittgenstein must have had in mind. If we approach set theory via the naïve comprehension principle, using examples to illustrate the principle in action – we have the set of all human mothers, the set of all mothers under thirty years old on 1 January 2000, the set of all teaspoons, and so on – with this background, Russell's Paradox comes as a surprise, even, as to Frege, a devastating bolt out of the blue. On the other hand we may prove in a couple of lines by *reductio* that there is no collection C of objects, no relation R on C and no object a of C such that for all x in C, xRa if and only if not xRx – all we need to do is to select $x = a$. From *this* elementary and general perspective, Russell's result follows unsurprisingly as a mere instance by setting C to sets and R to $\in$. Russell's famous barber example is just another instance. As Wittgenstein puts it, "If you are surprised, you have not understood it yet." (ibid.) *Post hoc*, the diagonalizing move that Russell makes, following the pattern set by Cantor, is a commonplace in logic and mathematics, to such an extent that we now find it surprising that Frege should *not* have noticed his logic with *Wertverläufe*[2] violated Cantor's proof that there are more subsets of any set than members of it. In this respect, Wittgenstein is surely right to say that once you see how things work, the surprise fades. Residual surprise is evidence of a lack of understanding, appreciation or firm grasp of how the proof works. Wittgenstein gives the example of being

[2] Value ranges, a kind of object associated with functions, extrapolate from the notion of the extensions of concepts (which are a kind of function for Frege) to all functions.

surprised by an unexpected reduction of a complex algebraic expression, and points out that the psychological effect of surprise is perhaps attendant on concentrating too much on the beginning and the end, and not enough on the steps in between. Surprise at a mathematical result cannot have the *mathematics* as its source: "The surprise and the interest [...] come, so to speak, from outside. I mean, one can say 'This mathematical investigation is of great psychological interest' or 'of great physical interest.'" (112)

Here is an example of how easy it is to misplace the source of surprise. In a lottery game, six numbers are selected at random from 49. One week, the draw throws up six consecutive numbers. "That's amazing!", says A. "No it's not!", says B, "those six numbers had just the same chance of coming up as any other six." B is right about this: in a fair lottery, every selection of six numbers is as likely to come up as every other. But A is right to be surprised. Only one in 317,814 combinations of 6 from 49 has six consecutive numbers, so on average such a combination would turn up, at a rate of two draws a week, about once every three thousand years. The surprise then is that it should happen in a short interval when we are taking note, that an event of such low probability should take place in such a short interval, and the source is physical. But B can rightly retort that any such distribution is equally probable, so the source of surprise is also psychological, since a distribution of six consecutive numbers is much more psychologically salient than all the other equiprobable distributions. In neither case does the mathematics *contribute* to the surprise: on the contrary, it helps to *explain* it.

Is there ontological surprise in pure mathematics?

Note that Wittgenstein's claim that surprise in mathematics always has a source outside the mathematics, in our own epistemic or imaginative limitations, or in something physical, works only for pure mathematics. There is a different issue about surprise, namely surprise at why and how concepts developed for purely mathematical purposes turn out to have unexpected, indeed startling applications in the physical world. For example, prime numbers, Hardy's favourite example of useless pure mathematics, not only serve an important auxiliary role in Gödel's surprising incompleteness theorems, but extending Fermat's Little Theorem about primes the mathematicians Ron Rivest, Adi Shamir, and Leonard Adleman devised the RSA algorithm for encrypting financial transactions on the internet. Complex Hilbert spaces are the formalism of choice for representing quantum mechanics, yet were developed solely for their own sake. Eugene Wigner (rather histrionically) called this "the unreasonable effectiveness of mathematics in natural science" (WIGNER, 1960), and it has been made the basis of Mark Steiner's account of the universe as "user-friendly" (STEINER, 1998). I must stress that while this is an interesting

debate in the area of philosophy of science, cosmology, and *applied* mathematics, it has a quite different point from Wittgenstein's: he was concerned only with the ultimate illegitimacy of surprise within pure mathematics itself.

Wittgenstein is concerned to dispel the idea that there is some *mystery* about pure mathematics, or that there is something deep and hidden, which surprises show up. In this I believe he is right. And his point itself is also neither deep nor mysterious. If pure mathematics consists in drawing conclusions from hypotheses by logically valid reasoning, then the reasons why people are surprised lie in their limitations: a proof is too long to keep all its steps in mind, so something from it is lost from an individual's view. Someone with a clear view of the whole proof from beginning to end will see it all as plain, each step following logically from its predecessors. It may be ingenious and wonderful, and the qualities of the author of the proof may inspire admiration and sometimes surprise, but the mathematics itself gives no legitimate ground for surprise:

> The demonstration has a surprising result!'—If you are surprised, then you have not understood it yet. For surprise is not legitimate here, as it is with the issue of an experiment. There—I should like to say—it is permissible to yield to its charm; but not when the surprise comes to you at the end of a chain of inference. For here it is only a sign that unclarity or some misunderstanding still reigns (111).

Limitations of memory or perception or of grasping complex propositions – in general, *epistemic* limitations – mean that a long or complex proof will be difficult to survey even for the adept. A putative derivation which is too long for anyone possibly to come close to appreciating, and which stubbornly resists such understanding, would cause the putative result simply to be set aside as not proven unless there were good evidence from other sources, such as computer testing, which gave other good reasons (not necessarily themselves pure mathematical reasons) for believing the result. Where the steps of a proof are followed one by one or in groups and found to be valid, but the overall structure remains elusive, the proof will be accepted as difficult and efforts made to understand the structure better or find a shorter proof, both of which are kinds of advance occurring many times in the history of mathematics.

If we hypothetically consider a mathematical proposition which followed logically from accepted hypotheses but whose proof could not possibly be made short enough for a finite creature to follow or appreciate or even write a computer program to test, then such a proposition will simply forever be left as undecided. There is no necessity that famous unresolved propositions about infinite domains such as Goldbach's Conjecture or Riemann's Hypothesis should be resolved at some time in the future, even if as a matter of logic they do (or their negations do) follow from the accepted assumptions.

There will probably always be a stream of pure mathematical results which even the most informed find initially surprising, because, as Wittgenstein rightly points out, until someone has worked through the proof and "looked around", they will know only the result and the starting point. But once the proof has been worked through and understood, the result will fall into place.

The limited role of surprise in pure mathematics can then be explained wholly in terms of the epistemic limitations of human beings in general and (even of) mathematicians in particular, in their difficulty in seeing how one proposition follows from others. There is no reason to call the existence of extra-mental mathematical reality into play to account for such surprises. In this, Wittgenstein was surely right, even if he understandably went too far in the other direction in trying to undermine the credentials of such initially surprising and even dismaying results as Gödel incompleteness.

References

LUKASIEWICZ, J. 'In Defence of Logistic', in his *Selected Works*. Amsterdam: North-Holland, 1970, p. 236–249.

SMITH, A. 'The Principles Which Lead And Direct Philosophical Enquiries: Illustrated By The History Of Astronomy', from his *Essays on Philosophical Subjects*, 1795, in *The Early Writings of Adam Smith*, ed. J. R. Lindgren. New York: Kelley, 1967, p. 22–35.

STEINER, M. *The Applicability of Mathematics as a Philosophical Problem.* Cambridge, Mass: Harvard University Press, 1998.

WIGNER, E. "The Unreasonable Effectiveness of Mathematics in the Natural Sciences", in *Communications in Pure and Applied Mathematics*, v. 13, 1960.

WITTGENSTEIN, L. 1978. *Remarks on the Foundations of Mathematics*. 3rd. Revised Ed. Oxford: Blackwell.

ARGUMENTOS
Revista de Filosofia

Axel Barceló*

Las tablas de verdad como filosofia

RESÚMEN

Pese a su aparente simplicidad, el método tradicional de tablas de verdad presupone un gran número de tesis filosóficas sobre la lógica clásica, de tal manera que es posible cambiar alguna de sus características – por ejemplo, el número de renglones – rechazando alguna de dichas presuposiciones. Como ilustración, en este artículo muestro cómo, si introducimos un tercero valor de verdad, el número de renglones aumenta; mientras que si las proposiciones con las que interpretamos las variables proposicionales de la fórmula no son lógicamente independientes entre sí, el número de renglones disminuye.

Palabras-clave: logicas multi-valuadas; lógicas intensionales; lógica modal; futuros contingentes; lógicas rivales.

RESUMO

Apesar de sua aparente simplicidade, o método tradicional de tabelas de verdade pressupõe um grande número de teses filosóficas sobre a lógica clássica, de tal maneira que é possível modificarmos algumas de suas características, por exemplo, o número de regras, rechaçando alguma de suas ditas pressuposições. Como ilustração, neste artigo mostro como, se introduzirmos um terceiro valor de verade, o número de regras aumenta; enquanto que, se as proposições com as quais interpretamos as variáveis proposicionais da fórmula forem logicamente independentes entre si, o número de regras diminuem.

Palavras-chave: lógicas multi-valoradas; lógicas intensionais; lógica modal; futuros contingentes; lógicas rivais.

* Professor at the Universidad Nacional Autónoma de México (UNAM), abarcelo@filosoficas.unam.mx

Las tablas de verdad son, por una parte, uno de los métodos más sencillos y conocidos de la lógica formal, pero la mismo tiempo también uno de los más poderosos y claros. Entender bien las tablas de verdad es, en gran medida, entender bien a la lógica formal misma. En este breve texto quiero ensayar algunas reflexiones alrededor de las tablas de verdad y, en particular, defender la idea de que no son herramientas lógicas sin contenido teórico, sino por el contrario, materializan varios principios lógicos que no sólo son controvertidas, sino que han sido cuestionados, dando pie a varias extensiones y rivales de la lógica clásica, las cuales pueden entenderse mejor cuando son concebidas, a su vez,como dando pie a extensiones y rivales de las tablas de verdad clásicas.

¿Qué es una tabla de verdad?

Fundamentalmente, una tabla de verdad es un dispositivo para demostrar ciertas propiedades lógicas y semánticas de enunciados del lenguaje natural o de fórmulas del lenguaje del cálculo proposicional:[1]

1. Sin son tautológicas, contradictorias o contingentes
2. Cuáles son sus condiciones de verdad[2]
3. Cuál es su rol inferencial, es decir, cuáles son sus conclusiones lógicas y de qué otras proposiciones se siguen lógicamente

La creación de este método suele atribuirse a Ludwig Wittgenstein (1921), y aunque ya era conocido en la tradición lógica-algebraica, y que Peirce (en notas no publicadas, anteriores a 1910[3]) y Post (1920) habían utilizado ya tablas de verdad, fueron Russell (1918) y Wittgenstein los que divulgaron este método como instrumento de análisis lógico-semántico en términos de condiciones de verdad.

[1] Existe, dentro de la filosofía de la lógica toda una discusión acerca de cuál de estos dos papeles es lemas importante. Hunter (1971), por ejemplo, sostiene que la verdad lógica es más fundamental que la validez, aunque la presentación clásica de esta posición se encuentra en (QUINE, 1969). Más recientemente, Ian Hacking (1979) y John Etchemendy (1990) han defendido la posición contraria: que la noción de consecuencia lógica es más fundamental. Es interesante notar que aquellos que piensan que la verdad lógica es más fundamental que la validez tienden a sostener que los métodos semánticos de análisis lógico, como las tablas de verdad, son superiores a los sintácticos, es decir, en términos de axiomas y reglas de inferencias, mientras que sus oponentes suelen tomar la posición opuestamente contrario. Hago una breve introducción a esta discusión en (BARCELÓ, 2014).

[2] Tradicionalmente, el método de tablas de verdad tiene como función para explicar cómo las condiciones de verdad de ciertos enunciados – aquellos que conocemos comúnmente como enunciados moleculares, es decir, enunciados complejos formados a partir de enunciados mas simples a través del uso de conectivas lógicas como "y", "o", etc. –, en función de las condiciones de verdad de sus componentes. Como tal, el uso de tablas de verdad para el análisis semántico de enunciados cabe de lleno dentro de la tradición composicionalista de análisis semántico según la cual (por lo menos ciertos aspectos importantes de) el contenido de enunciados complejos está determinado por el contenido de sus componentes y la forma en que éstos lo componen.

[3] Cf. Fisch and Turquette (1966) y Anellis (1994). Nótese que lo importante para la fundamentación del método semántico, no es lo que Shosky (1997) llama el dispositivo [*device*] de tablas de verdad, sino el método [*technique*] de tablas de verdad, es decir, el método de análisis veritativo-funcional del significado.

El procedimiento para construir una tabla de verdad es sencillo y relativamente mecánico; en este breve texto, asumiré que los lectores saben ya cómo hacer una tabla de verdad para cualquier fórmula del cálculo proposicional clásico: Para aplicar el método de tablas de verdad a un enunciado o proposición, recordemos, es necesario primero simbolizarlo, es decir, determinar qué fórmula del lenguaje proposicional muestra su forma lógica y, luego, elaborar la tabla de verdad de dicha formula. Si al aplicar el método de tablas de verdad encontramos que una fórmula es tautológica, presumimos que ella es una verdad lógica del cálculo proposicional es decir que es lógicamente válida, lógicamente verdadera o verdadera con necesidad lógica. Por lo tanto, el uso de las tablas de verdad como métodos para demostrar que algo es lógicamente necesario presupone ciertas tesis sobre la verdad y la necesidad lógicas. Cada uno de los pasos y cada una de las características de las tablas de verdad representa una tesis lógica sustancial.

Tomemos por ejemplo, el popular principio de que toda tabla tiene 2^n renglones, dónde la *n* corresponde al número de variables proposicionales (también conocidas como "letras proposicionales") que aparecen en la fórmula. Una fórmula de 3 variables proposicionales, por ejemplo, tendría $2^3=8$ renglones. Pero ¿por qué es esto? La respuesta más directa es que ése es el número de combinaciones que existen de asignaciones de valores de verdad a cada una de las variables. En otras palabras, porque si asignamos a cada variable uno de los dos valores de verdad – verdadero o falso –, las posibles combinaciones son exactamente ocho, ni más, ni menos. Si bien es una verdad matemática indudable que la combinatoria de dos valores a *n* número de variables es 2^n, para que este principio valga como principio lógico dentro de una demostración lógica – que, a fin de cuentas es lo que una tabla de verdad es –, es necesario que ciertas cosas sean verdaderas: Por ejemplo, entre otras cosas, es necesario que para determinar que una fórmula sea tautológica baste tomar en cuenta sólo cuál es el posible valor de verdad que tome la interpretación de sus variables proposicionales. También es necesario que se requieran considerar todas las posibles interpretaciones de las variables. Además, es necesario que a cada asignación de valores a las variables les corresponda uno y sólo un renglón. También es necesario que los valores de verdad sean dos – verdadero o falso. Si los valores de verdad fueran más, o fueran menos, las combinaciones posibles serían otras: más renglones si son más valores, y menos renglones si fueran menos valores. Además, el número de renglones a considerar también cambiaría si en cada renglón cada variable proposicional pudiera tener, no un sólo valor determinado, sino dos (o más) o ninguno. En este breve ensayo veremos brevemente no solamente qué sucede cuando algunas de estas cosas cambian, sino que también veremos qué razones tendríamos para pensar que deberíamos cambiarlas.

El primer principio que pondremos en cuestión es precisamente el principio de que la interpretación de toda variable proposicional no puede tener sino uno de los dos valores de verdad: verdadero y falso. A este principio se le conoce comúnmente como bivalencia y junto con el principio de no-contradicción ha sido considerado uno de los principios lógicos básicos. Se le llama también un principio *semántico* porque tiene que ver con la interpretación de los símbolos, es decir, con su significado. Sin embargo, no tiene que ver con ninguna interpretación o significado particular, sino con cualquier interpretación posible. Por eso es que sigue siendo un principio lógico y formal. Ahora bien, para poder entender el principio, por lo tanto, debemos entender también cómo se interpretan las variables proposicionales, a lo que dedicaremos la siguiente sección.

¿Qué es interpretar?

"Interpretar" significa asignar significados. En este sentido, es más o menos el proceso inverso a la simbolización o formalización que aprendemos en nuestros cursos básicos de lógica. En ellos aprendemos a traducir enunciados en fórmulas, es decir, a pasar del lenguaje ordinario y natural al lenguaje artificial de las fórmulas lógicas. Ahora bien, la interpretación es dar el paso inverso: asignar a cada fórmula una proposición.

Como su nombre lo indica, las variables proposicionales se interpretan por proposiciones. Interpretar una variable de este tipo es asignarle una proposición. A cada enunciado declarativo simple (es decir, que no está compuesto por otros enunciados, aunque él mismo sí sea parte de otros enunciados complejos) lo simbolizamos por una variable proposicional de tal manera que si dos enunciados significan lo mismo, es decir, si tiene como contenido la misma proposición, los simbolizábamos con la misma variable, esto es, con la misma letra. Así, cada variable proposicional simbolizaba una proposición.[4]

En consecuencia, cuando hablamos de las posibles interpretaciones de las variables proposicionales no hacemos sino hablar de las posibles proposiciones que se pueden simbolizar por variable de este tipo, es decir, todas. De tal manera que cuando decimos que todas las posibles interpretaciones de las variables proposicionales no pueden ser sino verdaderas o falsas – que es lo que dice el principio de bivalencia – lo que estamos diciendo es que *todas las proposiciones no pueden ser sino verdaderas o falsas.*

El principio de bivalencia ha sido tomado tradicionalmente como un principio lógico fundamental: toda proposición es verdadera o falsa. Si no es

[4] Este también es un principio implícito en la construcción de tablas de verdad que ha sido cuestionado (RUSSELL, 2008).

verdadera, es falsa y si no es falsa, es verdadera. No hay tercera opción. Por eso se le conoce también como principio del tercer excluso. La carga de la prueba descansa sobre quién defienda la tesis de que el principio es falso, es decir que existen más valores además de los dos tradicionales. Quién quiera defender la existencia de un tercer valor de verdad (o de otros más) tendría que mostrar:

1. Cuál sería ese tercer valor
2. En qué sentido es un valor de verdad
3. A qué (tipo de) proposiciones se le aplicaría, i.e., mostrar ejemplos de proposiciones que claramente no sean verdaderos ni falsos.
4. Cómo se comportarían lógicamente dichas proposiciones. Cómo interactuarían con otras proposiciones. es decir, cuál es su lógica. En particular, cómo afectaría la introducción de este nuevo valor nuestras tablas de verdad.

Como ejemplo, quiero hablar un poco de la primera lógica multivaluada, la cual toma como ejemplos paradigmáticos de enunciados que expresan proposiciones que no son verdaderas ni falsas a los futuros contingentes, es decir enunciados que refieren a hechos futuros que no son necesarios, sino que pueden darse o no de manera contingente. El ejemplo clásico, que le debemos a Aristóteles, es:

(A) "Mañana habrá una batalla naval"

Según defensores de un tercer valor de verdad, como Jan Lukazewicz (Mijangos 2003) – los que de ahora en adelante llamaremos "trivalentistas" –, si bien es cierto que, o bien mañana habrá una batalla naval o bien no la habrá, de ello no se sigue que la proposición que expresa el enunciado (A) sea verdadera o falsa. Si mañana hay una batalla naval, la proposición será verdadera y si no la hay, será falsa. Sin embargo, de ello solamente se sigue que mañana la proposición será verdadera o falsa. Pero esto no nos dice nada sobre hoy. Mas bien parece que hoy la proposición no es todavía ni verdadera ni falsa. Para que sea verdadera es necesario que mañana haya una batalla naval. Para que sea falso, es necesario que mañana no haya una batalla naval. Hasta mañana, no se cumplirán ninguna de las condiciones. La proposición, por lo tanto, por ahora carece de cualquiera de esos valores de verdad hoy. Ya mañana tendrá alguno. Hace poco más de una década, John MacFarlane (2003) desarrolló un nuevo argumento contra la bivalencia de los futuros contingentes. Según él, cuando mañana diga "Lo que dijiste ayer (es decir, que habría una batalla naval) es cierto" no estaré diciendo que la proposición era verdadera ayer, sino que eso que dijiste ayer es verdadero hoy.

Tablas de verdad trivalentes

Recordemos que nuestras tablas de verdad tradicionales pueden rescribirse si permitimos dejar vacías casillas en las que el valor de verdad de

la fórmula atómica es irrelevante, por ejemplo, podemos re-escribir así la tabla de la disyunción:

P	Q	PvQ
V		V
	V	V
F	F	F

Las primeras dos líneas señalan que no importa cuál sea el valor de verdad de uno de los disyuntos, siempre que el otro sea verdadero, la disyunción será verdadera. De la misma manera, podríamos abreviar la tabla de la conjunción de la siguiente manera:

P	Q	P&Q
V	V	V
	F	F
F		F

Las últimas dos líneas señalan que no importa cuál sea el valor de verdad de uno de los disyuntos, siempre que el otro sea falso, la conjunción será falsa.

La ventaja de este tipo de tablas para nuestros propósitos es que permiten extenderse de manera muy natural para permitir un tercer valor de verdad que no sea ni verdadera ni falso. Llamémosle "I" por "indeterminado". Ahora podemos usar nuestra tabla abreviada de la disyunción clásica para desarrollar una tabla de verdad (no abreviada) para la disyunción trivalente.

Primer paso: identificar las diferentes nueve posibilidades de combinaciones para dos variables:

P	Q	PvQ
V	V	
V	I	
V	F	
I	V	
I	I	
I	F	
F	V	
F	I	
F	F	

Segundo paso: Usamos las primeras dos líneas de la tabla abreviada para determinar el valor de verdad de los renglones con por lo menos un argumento verdadero:

P	Q	PvQ
V	V	V
V	I	V
V	F	V
I	V	V
I	I	
I	F	
F	V	V
F	I	
F	F	

Tercer paso: Cómo la última línea de la tabla abreviada es también la última línea de la nueva tabla, le corresponde el mismo valor de verdad: falso.

P	Q	PvQ
V	V	V
V	I	V
V	F	V
I	V	V
I	I	
I	F	
F	V	V
F	I	
F	F	F

Cuarto paso: Finalmente, cómo ya tenemos los renglones que son verdaderos o falsos según la tabla original, los renglones que aún no tienen valor de verdad, dado que no son ni verdaderos (sino hubieran quedado como tales en el segundo paso) ni falsos (ya que tampoco quedaron así en el tercer paso), deben ser indeterminados!

P	Q	PvQ
V	V	V
V	I	V
V	F	V

I	V	V
I	I	I
I	F	I
F	V	V
F	I	I
F	F	F

En algunos casos, esta tabla de verdad aparece, no en tres columnas, sino en un cuadro así:

v	V	I	F
V	V	V	V
I	V	I	I
F	V	I	F

Lo cual tiene la ventaja de dejar más claro el patrón que emerge de la tabla.

Si seguimos los mismos pasos para la conjunción, obtenemos las siguiente tablas:

P	Q	P&Q
V	V	V
V	I	I
V	F	F
I	V	I
I	I	I
I	F	F
F	V	F
F	I	F
F	F	F

&	V	I	F
V	V	I	F
I	I	I	F
F	F	F	F

Si comparamos las dos tablas cuadradas, podemos ver la simetría entre la conjunción y la disyunción.

¡Así, ya tenemos tablas de verdad con más de 2^n renglones! Además, una vez que entendemos qué sucede cuándo se introduce un nuevo valor de verdad, podemos imaginar cómo serían lógicas de cuatro o más valores de verdad. Es más, como Lukaciewicz y Boole (1854) mostraron ya hace más de un siglo, podemos fácilmente hablar de lógicas con un infinito de valores de verdad.

Tablas de verdad intensionales e independencia lógica

En clases básicas de lógica solemos aprender que una tabla de verdad tiene siempre 2^n renglones, dónde *n* es el número de ocurrencias de operadores lógicos en la fórmula o argumento que se esté simbolizando. Lo que comúnmente no se nos enseña es que, como bien señalo Wittgenstein ya en su *Tractatus Logico-Philosophicus* para que esto sea verdad, las variables deber simbolizar proposiciones atómicas o, por lo menos, lógicamente independientes entre sí (es decir, cada proposición simbolizada debe ser lógicamente independiente de las demás).

Para verificar que efectivamente estamos tratando con dos proposiciones independientes, A y B, es necesario que estas satisfagan cinco condiciones:

1. A no debe seguirse de B, es decir, debe ser posible que A sea verdadero y B falso
2. Y vice versa, B no debe seguirse de A, es decir, debe ser posible que B sea verdadero y A falso
3. La verdad de A debe ser compatible con la verdad de B, debe ser posible que tanto A como B sean ambos verdaderos *al mismo tiempo*, es decir, en la misma circunstancia.
4. La falsedad de A debe ser compatible con la de B, debe ser posible que tanto A como B sean ambos falsos *al mismo tiempo*, es decir, en la misma circunstancia.
5. Cuando sólamente tenemos una proposición, ésta no debe ser necesariamente verdadera ni necesariamente falsa.

Si no se cumplen alguna de estas condiciones, entonces alguna de los renglones posibles de la tabla representara como posible un caso que no es realmente posible. Si A se sigue lógicamente de B, por ejemplo, entonces ya no es posible que A sea verdadera y B falsa. Por ello, el renglón que le asigna verdadero a A y falso a B no representa una posibilidad real. Es necesario, por lo tanto, eliminarlo de la tabla.

Supongamos que queremos hacer la tabla de verdad del siguiente enunciado:

(2) Si tu hermano no hace el examen, no lo pasará.

Identificamos las proposiciones atómicas y les asignamos una variable:

P: Tu hermano hace el examen.
Q: Tu hermano pasará el examen.

De esta manera, podemos formalizar (2) como $(\sim P)\Rightarrow(\sim Q)$ y construir su tabla de la siguiente manera:

P	Q	(~*P*)®(~*Q*)
V	V	V
V	F	V
F	V	F
F	F	V

Sin embargo, hay algo extraño en el análisis que presenta esta tabla, ya que nos dice, entre otras cosas, que el enunciado sería falso si *P* fuera falso y *Q* verdadero, es decir, si tu hermano no hiciera el examen y, sin embargo, lo pasará, ¡lo cual es imposible! Por eso es que pareciera que este renglón no debería de aparecer en la tabla, ya que no es una posibilidad sino una imposibilidad. Así pues, la tabla de verdad correcta debería ser algo así cómo:

P	Q	(~*P*)®(~*Q*)
V	V	V
V	F	V
F	F	V

Y ahora sí podemos ver que, en realidad, ¡el enunciado expresaba una tautología! Desde esta perspectiva, por lo tanto, las fórmulas no son tautológicas, contradictorias o contingentes *en sí mismas*, sino *en una tabla*, y qué tabla sea la adecuada para evaluar una formula no va a depender de la fórmula misma, sino de su interpretación, es decir, de qué proposiciones simboliza cada variable proposicional. Por ello, mucha gente dice que este tipo de tablas no respetan el principio según el cual las propiedades lógicas de una proposición, en particular si una proposición es tautológica o no, debe depender sólo de su forma, no de su interpretación particular.

El que una fórmula sea tautológica, contradictoria o contingente, depende por supuesto, de cuales son los renglones de la tabla en la que se evalúa. La misma fórmula puede ser contingente en una tabla, contradictoria en otra y tautológica en otra más, dependiendo de qué renglones tenga la tabla en cuestión. Hay formulas que siempre serán tautológicas o contradictorias, no importa en qué tablas las evaluemos. Estas son las tautologías y contradicciones que ya conocemos de nuestro cálculo proposicional. En otras palabras, si una fórmula es tautológica en la tabla de verdad tradicional de 2^n renglones, entonces será tautológica en cualquier otra tabla de verdad. Si una formula es verdadera en todos los renglones, no importa qué renglones eliminamos, seguirá siendo

verdadera en todos ellos. Lo mismo sucede con las formulas que resultan contradictorias en las tablas de 2^n renglones: también son contradictorias en cualquier otra tabla. Por el contrario, si una fórmula es contingente en la tabla de 2^n renglones, entonces dependerá de qué renglones se incluyan o eliminen de la tabla para que sea contradictoria, tautológica o contingente.

La área de la lógica que estudia las propiedades y relaciones lógicas expresadas en este tipo de tablas se le llaman lógicas *intensionales*, y el trabajo fundamental se lo debemos a Rudolf Carnap (1947), aunque suelen estudiarse dentro del marco semántico introducido por Saul Kripke (1963) en sus estudios sobre la modalidad.[5] A decir verdad, en ningún libro de texto de lógica intensional encontrarán nunca una tabla de verdad (recortada). Por el contrario, las lógicas intencionales suelen introducirse apelando a la noción de *mundo posible,* el concepto fundamental de la semántica intensional. Sin embargo, esto no debe confundirnos. Las nociones semánticas de mundo posible y de tablas de verdad están íntimamente ligadas ya que los renglones de una tabla de verdad no representan otra cosa sino tipos de mundos posibles y vice-versa. Recordemos que cada renglón de la tabla representa una posible manera de ser las cosas. Normalmente, solamente una de ellas es la manera cómo las cosas realmente son, digamos, en el mundo real. Los demás renglones representan las manera en que las cosas podrían ser, pero de hecho no son, es decir, las manera en que las cosas podrían ser en otros mundos meramente posibles.

Pongamos un ejemplo. Supongamos que queremos hacer la tabla de verdad del siguiente enunciado:

(3) Si tu hermana no pasa el examen, estarás en graves problemas.

Identificamos las proposiciones atómicas y les asignamos una variable:

P: Tu hermana pasa el examen

Q: Estarás en graves problemas

De esta manera, podemos formalizar (3) como $(\sim P)\Rightarrow Q$ y construir su tabla de la siguiente manera:

P	Q	$(\sim P)\circledR Q$
V	V	V
V	F	V
F	V	V
F	F	F

[5] Antes de Kripke, C. I. Lewis (1914) había hecho ya trabajo sustancial en esta dirección, pero su trabajo era sintáctico, es decir, no tenia nada que ver con el tipo de análisis semántico que se lleva a cabo con tablas de verdad.

¿Qué es lo que nos dice el primer renglón de la tabla? Nos dice que si P y Q son ambos verdaderos, (~P)⇒Q también lo es. En otras palabras, en toda circunstancia o mundo posible en que "Tu hermana pasa el examen" y "Estarás en graves problemas" sean verdaderos, será una en que "Si tu hermana no pasa el examen, estarás en graves problemas" también será verdadero. Es decir, todo mundo posible en el que tu hermana pasa el examen y estarás en graves problemas es un mundo en el que, si tu hermana no pasa el examen, estarás en graves problemas.

Uno podría pensar que una diferencia importante entre la manera tradicional de hacer semántica intensional en términos de mundos posibles y usando tablas de verdad es que, en la manera tradicional solemos distinguir uno, entre los mundos posibles, como el mundo real. Sin embargo, esto podría hacerse fácilmente añadiendo una convención para distinguir entre los renglones de la tabla, uno como correspondiendo a como son las cosas en realidad. Por ejemplo, si efectivamente tu hermana pasa el examen pero no estarás en problemas, podríamos añadir esta información a la tabla de verdad marcando de alguna manera el renglón correspondiente en la tabla, por ejemplo, así:

P	Q	(~*P*)®*Q*
V	V	V
V	F	V
F	V	V
F	F	F

Así, la misma tabla nos diría no sólo qué valores tendría el enunciado bajo en análisis en diferentes circunstancias posibles, sino que también nos diría qué valor de verdad tiene en el mundo real (en este caso, es verdadero).

Ahora bien, ¿qué sucede con la relación de accesibilidad, fundamental en las semánticas de mundos posibles tipo Kripke? Recordemos que, en las semánticas de mundos posibles tradicionales es posible expresar que el que una circunstancia o mundo sea realmente posible depende de como el mundo realmente es (o podría ser), y esto se logra a través de la noción de *accesibilidad,* de tal manera que un mundo *w* es posible si el mundo *x* es real si y sólo si el mundo *w* es accesible desde el mundo *x*. La idea de fondo, una vez más, es que no toda circunstancia de evaluación que podamos representar corresponde o puede corresponder a una posibilidad genuina. Pero ya vimos que esto lo podemos representar en el método de tablas de verdad precisamente eliminando los renglones que no correspondan a posibilidades genuinas.

En este respecto, la única ventaja real que ofrecen la manera tradicional de representar las semánticas tipo Kripke es que permite distinguir, no sólo entre lo que es posible y lo que no es, sino que también entre lo que podría ser posible y lo que no podría, entre lo que podría poder ser posible y lo que no, etc. En otras palabras, no sólo te permite representar lo que sería posible si algo fuera el caso, sino también como podrían ser las cosas si alguna de esas posibilidades se actualizara, o si alguna de estas nuevas posibilidades se actualizara, y así ir repitiendo el mismo proceso de manera recursiva tanto como uno quiera.

Uno podría bien pensar que, en realidad, esto podría hacerse también cambiando las tablas de verdad, de tal manera que pudiéramos asociar a cada renglón de la tabla de verdad otra tabla de verdad que representara las posibilidades genuinas correspondientes a dicha posibilidad, y luego otras tablas mas a cada uno de los renglones de cada una de esas tablas y así *ad libitum*. Sin embargo, aunque dicha estrategia efectivamente funcionará, al hacerlo no estaríamos haciendo más que incorporando la noción de accesibilidad a nuestro método de tablas de verdad y, en efecto, estaríamos diluyendo casi por completo la manera tradicional de presentar las semánticas Kripkeanas y nuestro nuevo método de tablas de verdad. En otras palabras, estaríamos mostrando como, detrás de la manera tradicional de hacer semántica intensional, siguen sobreviviendo las intuiciones básicas del método de tablas de verdad.

Otras tablas de verdad divergentes

Además de las tablas polivalentes e intensionales, hay muchas otras tablas de verdad *raras*, de las cuales no hablaré aquí, peor no quiero dejar de mencionar. Por ejemplo, hay tablas de verdad en las que los renglones se bifurcan en dos o más sub-renglones y son útiles para lo que en lógica llamamos *super-valuaciones.* También existen tablas con *n* valores y más de *2n* renglones, ¿cómo es posible? Pues porque, a diferencia de las tablas tradicionales, en estas tablas el orden de los renglones *sí* importa, de tal manera que renglones repetidos cuentan como renglones distintos. Finalmente, también existen las tablas *bidimensionales*, usadas originalmente en ciertas lógicas intensionales, pero popularizadas gracias al trabajo de Robert Stalnaker y otros (SCHROETER, 2012). Todas ellas extienden o rivalizan los principios lógicos y/o semánticos de la lógica clásica, dando pie a tablas de verdad distintas de las que estamos acostumbrados. Como espero haya quedado claro, el campo es muy amplio y en este texto apenas he rozado lo su superficie. Sin embargo, creo haber dicho lo suficiente para convencerlos de que hay mucha filosofía en una tabla de verdad.

Referencias bibliográficas

AGNELLIS, I. "The genesis of the Turth-Table Device". *Russell*, v. 34, 2004. p. 55-70.

BARCELÓ, A. "Verdad Lógica: Enfoqes Sintáctics y Semánticos". In: DÍAZ HERRERA P. y J. JASSO MÉNDEZ. *Problemas contemporáneos de Filosofía*. México: Universidad de la Ciudad de México, 2014, p. 73-96.

BOOLE, G. *An Investigation of The Laws of Thought on which are founded the Mathematical Theories of Logic and Probabilities.* Londres*:* Macmillan, 1854.

CARNAP, R. *Meaning and Necessity.* Chicago: University of Chicago Press, 1947.

ETCHEMENDY, J. *The Concept of Logical Consequence.* Cambridge Mass: Harvard University Press, 1990.

FISCH, M. & A. TURQUETTE. "Peirce's Triadic Logic". *Transactions of the Charles S. Peirce Society* v. 2, n. 2, 1966, p. 71 - 85.

HACKING, I. What is Logic? *Journal of Philosophy, v.* 76, 6, 1979, p. 285-319.

HUNTER, G. *Metalogic:* an Introduction to the Metatheory of Standard First Order Logic. Berkeley: University of California Press, 1971.

KRIPKE, S. "Semantical considerations on modal logic". *Acta Philosophica Fennica*, v. 16, 1963, p. 83-94.

LEWIS, C. I. "The Calculus of Strict Implication". *Mind,* v. 23, 1914, p. 240–247.

MACFARLANE, J. "Future Contingents and Relative Truth? *The Philosophical Quarterly*, v. 53, n. 212, 2003, p. 321–336.

MIJANGOS, T. *Futuros Contingentes y Polivalencia:* La Propuesta de Jan Lukasiewicz*"*. Tesis de maestría en filosofía. Xalapa: Facultad de Filosofía de la Universidad Veracruzana, 2003.

POST, E. L. "Determination of all closed systems of truth tables". *Bulletin American Mathematical Society, v.* 26, 1920, p. 437.

QUINE, W. V. O. *The Philosophy of Logic.* Engelwood: Prentice-Hall, 1969.

RUSSELL, B. *The Philosophy of Logical Atomism.* London: Fontana, 1918.

RUSSELL, G."One true logic?" *Journal of Philosophical Logic,* v. 37, n. 6, 2008, p. 593–611.

SHOSKY, J. "Russell's Use of Truth Tables". *Russell*, v. 17, 1997, p. 11–26.

SCHROETER, L. "Two-Dimensional Semantics". En: ZALTA, E. *The Stanford Encyclopedia of Philosophy*. Disponível em: <http://plato.stanford.edu/archives/win2012/entries/two-dimensional-semantics/>.

WITTGENSTEIN, L. *Tractatus Lógico-Philosophicus.* Traducción de Jacobo Muñoz e Isidoro Reguera. México: Alianza Editorial, 1921.

ARGUMENTOS
Revista de Filosofia

Cícero Antônio Cavalcante Barroso*

Uma explicação cognitiva do 'segue-se'

RESUMO

O principal ponto de partida de qualquer lógica dedutiva é o fato de que alguns enunciados se seguem necessariamente de outros. A lógica fornece regras que nos permitem demonstrar essas conexões entre enunciados, mas ainda é possível indagar por que devemos aceitar essas regras. Há várias respostas possíveis. Neste artigo, farei uma rápida análise de algumas delas, mas me concentrarei em expor e analisar a resposta cognitiva, segundo a qual as regras da lógica devem ser aceitas porque temos certos mecanismos inatos de processamento dedutivo que nos capacitam a ver que as inferências lógicas elementares são válidas. Ao analisar essa explicação, tentarei mostrar também que ela parece nos remeter a uma tese metafísica mais forte, a saber, a tese de que o desenho de nosso módulo de processamento lógico nos diz algo sobre as propriedades lógicas do nosso mundo.

Palavras-chave: Inferência dedutiva; base neurobiológica da dedução; realismo lógico.

ABSTRACT

The main starting point of any deductive logic is the fact that some statements necessarily follow from others. The logic provides rules that allow us to demonstrate that connections between statements, but you can still ask why we should accept these rules. There are several possible responses. In this article, I will briefly analyze some of them, but I will focus on exposing and analyzing the cognitive response, by which the rules of logic should be accepted because we have certain innate mechanisms of deductive processing that enable us to see that the elementary logical inferences are valid. By analyzing that explanation, I will also try to show that it seems to lead to a stronger metaphysical thesis, *viz.* that the design of our module of logical processing tells us something about the logical properties of our world.

Keywords: Deductive inference; neurobiological basis of deduction; logical realism.

* Professor Adjunto da Universidade Federal do Ceará. Email: cicero@lia.ufc.br

O que a tartaruga de Carroll disse para Aquiles

O título deste artigo talvez seja mais permissivo do que deveria. Um título mais preciso seria: "Uma explicação cognitiva da inferência dedutiva". Com efeito, na linguagem ordinária, podemos usar a expressão "segue-se", ou outras equivalentes ('conclui-se', 'logo', 'portanto', 'assim', 'consequentemente' etc.), para expressar diferentes variedades de inferência, e não me interessa aqui tratar de todas elas. Meu interesse primário é a inferência dedutiva.

Quando falo de inferência dedutiva, penso naquele tipo de inferência que fazemos quando, apoiados apenas nos aspectos formais dos dados de entrada, estabelecemos que certo enunciado se segue necessariamente daqueles dados. Dito de outro modo, quando falo de inferência dedutiva, penso em silogismos, em instâncias da redução ao absurdo, em instâncias da generalização existencial, em inferências estabelecidas por provas lógicas complexas etc., e penso ainda em alguns raciocínios que não são totalmente codificados em uma linguagem proposicional, dependendo também de diagramas ou representações gráficas (por exemplo, os raciocínios que fazemos para preencher um quadro de sudoku).

Se minha argumentação fosse depender de uma distinção rigorosa entre inferência dedutiva e não dedutiva, talvez essa caracterização inicial da inferência dedutiva não fosse suficiente. Talvez fosse preciso esclarecer melhor o que são esses aspectos formais que nos permitem montar raciocínios dedutivos, e fosse preciso especificar em que sentido a conexão entre a conclusão e os dados de entrada é necessária. Mas, para meus propósitos neste artigo, basta supor que há casos de inferência que a maioria das pessoas com uma instrução básica em lógica concorda em chamar de 'inferência dedutiva' ou 'dedução'. Gostaria que me fosse permitido começar com essa suposição. Posto isso, fica acertado que o foco deste artigo estará neste tipo de inferência.

Há várias questões que podemos levantar em relação à dedução, e diferentes disciplinas do conhecimento tentam responder diferentes questões. Algumas questões egrégias são as seguintes: (i) De que forma os falantes fazem deduções em suas práticas linguísticas? (questão investigada pela Sociolinguística); (ii) Quais processos mentais são solicitados quando fazemos uma dedução? (questão que interessa a alguns campos das Ciências Cognitivas); (iii) Com base em quais princípios e regras nossas deduções devem ser estruturadas? (questão que a Lógica busca responder); (iv) Por que nossas deduções devem ser estruturadas do modo como a Lógica prescreve? (questão do âmbito da Filosofia da Lógica). Meu objetivo principal neste artigo é dar uma resposta para a questão (iv). Não obstante, a resposta que oferecerei busca respaldo nas respostas que as ciências cognitivas têm dado para a questão (ii). Desse modo, embora minha discussão de fundo seja eminentemente filosófica, minha argumentação faz um uso razoável de resultados empíricos oriundos do campo das

ciências cognitivas. É exatamente esse uso que caracteriza a minha explicação da inferência dedutiva como uma explicação cognitiva.

Talvez não seja claro para todos por que deveríamos nos preocupar com a questão (iv). Alguém pode expressar a opinião de que uma lógica é apenas um jogo no qual a única coisa que temos que fazer é seguir as regras que foram estabelecidas pelo lógico. As regras em si não precisam de justificação, da mesma forma como as regras do xadrez não precisam de justificação; ambas as modalidades de regras seriam como são por causa de uma decisão arbitrária, uma decisão do criador do xadrez no caso do xadrez, e uma decisão do lógico no caso da lógica. No entanto, embora essa opinião seja possível, ela não parece coerente com o trabalho efetivo do lógico que cria uma lógica dedutiva. Aparentemente, o que uma lógica dedutiva pretende estabelecer em primeiro lugar é um método que nos permita demonstrar que uma fórmula α qualquer se segue necessariamente do conjunto de fórmulas Γ, sempre que é verdade que α se segue necessariamente de Γ. Mas, se é essa a pretensão, a demonstração de que α se segue necessariamente de Γ deveria nos convencer desse fato, pois se um fato logicamente necessário é demonstrado, não podemos negá-lo nem em imaginação. Dessa forma, parece que uma regra lógica não pode ser do jeito que der na cabeça do lógico, ela precisa propiciar inferências convincentes. Nesse sentido, parece legítimo levantar a questão (iv).

Uma forma mais divertida de entender a pertinência da questão (iv) é introduzi-la *à la* Carroll. A estória foi apresentada inicialmente em Carroll 1985 e hoje é bem conhecida. Aquiles e a tartaruga estão conversando sobre argumentos lógicos e Aquiles dá um exemplo de um argumento logicamente válido: dadas as premissas A e B, a conclusão Z se segue necessariamente. Só que a tartaruga não se convence. Ela aceita as premissas, mas não aceita a conclusão. Aquiles explica que há uma regra que diz que se você tem A e B, você tem Z. Isso parece razoável para a tartaruga. Ela aceita essa regra, que chama de C, assim como continua aceitando A e B. O problema é que ela ainda não aceita Z. Aquiles faz nova tentativa propondo a regra D que diz que se você tem A, B e C, é necessário que você tenha Z. Mas a tartaruga está irredutível. Ela aceita A, B, C e D, mas não aceita Z. Logo fica claro que as regras propostas por Aquiles não terão o poder de convencer a tartaruga a aceitar Z nem em um milhão de anos. Em todo caso, a intuição que temos ao ler essa estória é a de que a tartaruga não poderia sinceramente repelir para sempre a conclusão. Pensamos que em algum momento ela deveria se convencer de Z porque o fato de ela estar sinceramente convencida das regras lógicas e das premissas é suficiente para lhe fazer ver que Z é irrecusável. Mas, se pensamos assim, é porque compreendemos que as regras lógicas não podem ser de qualquer jeito, elas devem ser tais que não seja possível aceitá-las e ao mesmo tempo recusar uma conclusão que elas impõem.

Dessa forma, a questão (iv) se qualifica como uma questão legítima. Uma resposta geral a ela seria esta: nossas deduções devem ser estruturadas do modo como a Lógica prescreve porque elas precisam ser convincentes. Mas essa resposta não é suficientemente esclarecedora. Há que se buscar uma resposta que esclareça também o que faz com que uma dedução feita com base em certa regra seja convincente. Há várias explicações que visam dar esse esclarecimento, e uma delas é a cognitiva. Nas seções seguintes vou tratar basicamente da explicação cognitiva, mas, neste final de seção, cumpre dizer algumas coisas sobre outras três explicações, a saber, as explicações semântica, sintática e sociolinguística.

A explicação semântica se caracteriza pela tese de que os preceitos da lógica são aceitáveis na medida em que eles nos garantem a validade das provas. Assim, nessa explicação, uma dedução convincente é antes de tudo uma inferência preservadora da verdade. E, sem dúvida, essa é uma qualidade indispensável em uma dedução. O mínimo que se pode exigir de uma inferência dedutiva é que ela nos permita constatar que, quando as premissas são verdadeiras, a conclusão também necessariamente o é.

O problema com essa explicação é que ela parece deixar algo em aberto. Pense no trabalho do lógico! Primeiro ele estabelece uma meta: "os argumentos produzidos com a minha lógica devem ser válidos". Depois ele trabalha para achar um conjunto de regras que lhe possibilite alcançar essa meta e, quando finalmente compõe tal conjunto, ele tenta provar que o teorema da correção é válido para ele. Ora, para provar o teorema da correção, o lógico precisa de uma definição da verdade, mas não há apenas uma. É um fato inegável que a definição de verdade que adotamos depende da interpretação que fazemos dos conectivos lógicos. Então o que fica em aberto é qual definição de verdade se deve escolher. Há uma mais natural? Mas o que a torna mais natural? Talvez alguém argumente dizendo que não importa qual definição de verdade o lógico adota, o importante é que ela é a definição adequada para as interpretações que ele faz dos conectivos. Isso é plausível, mas ainda não fica claro como podemos saber que uma definição é adequada e outra não. Quem no final das contas pode nos dizer o que é mais adequado em termos de semântica e, consequentemente, o que é mais adequado em termos de regras de inferência?

A explicação sintática tenta uma abordagem diferente para responder por que os preceitos da lógica são aceitáveis. Ela se apoia na tese de que, se aceitarmos tais preceitos, teremos garantia de que nossas deduções serão analisáveis em termos de operações lógicas intuitivas. Isso certamente deve fazer com que nossos argumentos dedutivos sejam convincentes. Aqui também vemos uma tentativa de justificar o aparato lógico com base no argumento de que ele dota nossas deduções com uma propriedade altamente desejável. Nesse caso, a propriedade almejada é o caráter analítico da prova, a propriedade que permite que uma inferência mais complexa possa ser quebrada em

inferências mais simples e imediatas. Mais uma vez, parece totalmente justificado que desejemos a propriedade e tentemos mostrar que os preceitos da lógica são igualmente desejáveis pelo fato de proverem nossas deduções com tal propriedade. Mas ainda aqui ficamos com a impressão de que não explicamos tudo o que deveria ser explicado.

Dessa vez, o que não está bem explicado é algo que está na base da ideia de análise: o conceito de inferência elementar. Com efeito, quando se afirma que deduções são analisáveis, o que se supõe é que a inferência dedutiva é uma construção que pode ter diferentes medidas de complexidade, e que uma inferência mais complexa deve poder ser construída a partir de inferências mais simples, e assim também com essas mais simples, até que alcancemos a medida mínima de complexidade. Nesse ponto, encontramos as inferências elementares. Essas inferências já não são analisáveis e o único modo de justificá-las é alegando que elas são intuitivas. E é exatamente aí que sentimos falta de uma explicação adicional. O que torna certas deduções intuitivas? Não pode ser simplesmente o fato sintático de que com elas podemos construir qualquer dedução mais complexa. Deduções não intuitivas também poderiam servir igualmente bem a esse propósito. Destarte, uma resposta em termos de propriedades sintáticas não parece suficiente.

Uma terceira possibilidade de explicação é a sociolinguística. Ela se baseia na tese de que: "Princípios de inferência dedutiva são justificados por sua conformidade com a prática dedutiva aceita" (GOODMAN, 1983, p. 63). De acordo com essa explicação, portanto, o que o lógico faz é verificar os modos como pragmaticamente raciocinamos e expressá-los de modo formal e sistemático. Isso parece estar de acordo com o fato de que as pessoas em geral já conseguem raciocinar antes de estudar lógica e que muitas vezes esses raciocínios são sancionados pela lógica. Isso não pode ser simplesmente uma coincidência. Isso deve acontecer porque os preceitos da lógica derivam de preceitos pragmáticos de raciocínio.

O problema com essa explicação, porém, é que ela não parece oferecer uma elucidação plausível do fato de que algumas de nossas práticas dedutivas aceitas não são sancionadas pela lógica clássica (um exemplo dado por Oswaldo Chateaubriand é a falácia da afirmação do consequente, que é amplamente usada na prática dedutiva comum, mas não recebe a aprovação da lógica (cf. CHATEAUBRIAND, 2001, p. 21). Talvez se possa dizer que o lógico não pode sancionar todas as práticas dedutivas aceitas porque assim ele não poderia construir um sistema lógico consistente. Isso, certamente é verdade, mas ainda é um fato evidente que o lógico sanciona algumas inferências e rejeita outras, e a explicação sociolinguística não parece ter nada de elucidativo a dizer sobre o que justifica essa segregação.

Em suma, quando tentamos responder à questão de por que nossas deduções devem ser estruturadas do modo como a Lógica prescreve, nem a ex-

plicação semântica, nem a explicação sintática, nem a explicação sociolinguística nos dão uma resposta totalmente satisfatória, embora possamos concordar que elas põem em relevo alguns pontos importantes da questão. Em vista disso, parece que estamos justificados em procurar uma explicação alternativa. A explicação alternativa que proponho que passemos a considerar a partir deste ponto é a explicação cognitiva.

O que a Ciência Cognitiva tem dito sobre a lógica

A explicação cognitiva para a aceitabilidade das regras da lógica se vale de certos resultados de pesquisas empreendidas em diferentes áreas das Ciências Cognitivas, especialmente na Linguística, na Psicologia e na Neurociência. Esses resultados nos fornecem uma base tanto teórica quanto experimental para afirmar que nossa capacidade de fazer deduções elementares não é aprendida, o que significa que possuímos certos mecanismos cognitivos que são biologicamente programados para computar essas deduções. A tese da explicação cognitiva é a de que os preceitos da lógica são aceitáveis porque nossos mecanismos inatos de processamento dedutivo podem reconhecê-los como válidos. Eles são, por assim dizer, programados para reconhecê-los.

O nativismo lógico tem origens bastante antigas, mas sua formulação contemporânea está ligada às pesquisas de Noam Chomsky no campo da Linguística. Nessas pesquisas, ele concluiu que nossa capacidade para desenvolver e dominar a linguagem é inata. Um fato foi essencial para que ele adotasse essa perspectiva sobre a linguagem, o fato de que nós aprendemos a nossa primeira língua sem que ninguém nos ensine. Antes da *gramática gerativa transformacional* (a teoria linguística chomskyana) a explicação mais aceita para esse fato era a de que, embora os bebês não tenham aulas para aprender sua primeira língua, eles observam como as pessoas se comunicam e aprendem por imitação e condicionamento. Chomsky notou que essa explicação era muito simplista e propôs então que os bebês já nascem com um tipo de teoria de linguagem e aprendem sua língua materna testando essa teoria. O fato de crianças de todos os lugares do mundo cometerem os mesmos tipos de erro mais ou menos na mesma época, enquanto estão aprendendo a falar, seria uma evidência disso. O fato de elas desenvolverem uma competência linguística mínima aproximadamente no mesmo espaço de tempo também seria uma indicação. A evidência principal, porém, é o fato de que uma língua natural é uma estrutura extremamente complexa. O estímulo linguístico que nós recebemos é muito pobre para explicar a exuberância da linguagem que nós apresentamos em poucos anos de vida. É preciso postular que a estrutura básica da linguagem está enraizada no nosso aparato cognitivo. Se não fosse pelo fato de possuirmos uma gramática embutida em nossas mentes, aprender uma língua seria uma tarefa absurda-

mente difícil ou mesmo impossível. É ancorado nessas evidências que Chomsky enuncia seu nativismo linguístico.

Ao postular que a estrutura básica da linguagem é inata, Chomsky está assumindo que a mente humana é um tipo de sistema computacional composto de diferentes programas interligados. Há programas responsáveis pela visão, pelas ações motoras, pelo raciocínio lógico e assim por diante. Em particular, há um sistema mental responsável pela linguagem. Chomsky o chama de "faculdade da linguagem". A faculdade da linguagem deve executar e gerenciar tarefas múltiplas tais como articulação e interpretação de fonemas, representação de aspectos semânticos, aplicação de regras de formação de expressões complexas etc. O fato de que nós humanos nascemos equipados com esse sistema é o que explica nossa capacidade de aprender e dominar uma língua. Em outras palavras, é isso que explica nossa competência e desempenho linguísticos.

A tese de que há uma faculdade da linguagem implica na afirmação de que há uma gramática universal, ou seja, há princípios que determinam as características fundamentais de todas as línguas naturais. A justificativa é a seguinte: se há estratégias de aprendizagem específicas para a aquisição da nossa primeira língua, isso é porque as rotinas envolvidas nessas estratégias devem nos predispor para assimilação da gramática da língua. Acontece que há uma grande diversidade de línguas naturais, mas a faculdade da linguagem deve ser a mesma em toda a espécie humana. Dessa forma, a mesma faculdade que possibilita que uma criança aprenda alemão, possibilita que outra criança aprenda swahili. Daí é possível concluir que todas essas línguas, aparentemente tão diferentes, em um nível mais profundo de análise, estão fundadas sobre os mesmos princípios, e são esses princípios que a faculdade da linguagem nos dá por antecipação. São esses princípios que constituem a gramática universal, o estado inicial da faculdade de linguagem. Tal gramática é, portanto, uma teoria que todo ser humano traz embutida na sua mente e que modela os modos que a linguagem pode assumir. Quando uma criança é exposta a uma língua natural, ela começa a testar sua teoria internalizada. À medida que os dados empíricos confirmam suas hipóteses, ela vai adquirindo domínio sobre a sua língua materna. De acordo com Chomsky, se não fosse assim, aprender a primeira língua seria uma tarefa virtualmente impossível.

Agora, um aspecto importante de nossa habilidade linguística é nossa capacidade de fazer relações inferenciais entre enunciados. De fato, essa capacidade parece central para a estruturação de nossa base linguística e para a nossa consequente competência conversacional. Muitas vezes, para entendermos o que os outros dizem, precisamos tirar conclusões a partir do que eles efetivamente afirmam e de outras coisas que já sabemos. Isso implica que a compreensão das palavras lógicas ('todo', 'ou', 'se', etc.) é fundamental para o nosso desenvolvimento linguístico. Como adquirimos o significado dessas pa-

lavras? Segundo Chomsky, não os adquirimos. Esses significados são fornecidos por nossa faculdade de linguagem, ou mais especificamente por um nível dessa faculdade, uma estrutura cognitiva que Chomsky chamou de *Logical Form* (LF). Crain & Khlentzos 2010 explicam que a LF funciona como uma definição de verdade. Em suas palavras:

> LF contains an interpreted vocabulary of logical words, including sentential connectives expressed in human languages by words like 'and' and 'or' and quantificational devices like 'every' and 'some'. For example, a logical expression corresponding to 'and', call it '&', can be found at LF. The semantic representation of '&' is such that a structure of the form [S & S'] will be true iff both S and S' are true (regardless of order). Given that Universal Grammar is the initial state of the language learner, the task of the child exposed to English is to figure out that the English word 'and' maps onto the LF expression '&'; the task of a child exposed to Japanese is to figure out that 'mo' maps onto '&', and so on. This view leads to the expectation that children learning any human language will 'know' the truth conditions of its logical words as soon as these words enter their speech. (CRAIN & KHLENTZOS, 2010, p. 31-32).

Uma consequência imediata da existência de uma faculdade inata como LF é a de que nossa habilidade para fazer e avaliar inferências dedutivas é inata. Com efeito, quando deduzimos um enunciado a partir de certas premissas, os aspectos formais das premissas que nos permitem fazer a inferência resultam da interpretação das palavras lógicas que essas premissas contêm. Uma vez que aceitemos que nossa interpretação dessas palavras é inata, temos que aceitar que nossa habilidade dedutiva também é inata.

Essa teoria da base inata de nossa capacidade dedutiva tem sido corroborada por vários experimentos realizados tanto por psicólogos como por neurocientistas. Uma estratégia de experimentação comum no campo da psicologia cognitiva é a aplicação de testes que visam determinar o modo como crianças pequenas entendem os conectivos lógicos. Vários desses testes têm mostrado que a interpretação que as crianças fazem das palavras lógicas é condizente com a hipótese da LF (CRAIN *et al*., 1996, CRAIN & KHLENTZOS 2010). É preciso notar, no entanto, que há controvérsias importantes em relação à interpretação do 'se' (JOHNSON-LAIRD & BYRNE, 2009).

Em paralelo com essa linha de investigação, tem se desenvolvido também a pesquisa a respeito da base neurofuncional da nossa habilidade dedutiva. Nesse domínio, a teoria hoje mais aceita é a Teoria do Processo Dual do Raciocínio (EVANS, 2003), segundo a qual temos dois módulos cognitivos distintos responsáveis por nossas habilidades inferenciais. Um desses módulos, que podemos chamar de MIBC (módulo de inferência baseada em crença), é especializado em inferências não dedutivas, inferências que dependem de nossas crenças a respeito do uso e da referência de termos não lógicos. O outro módulo, que podemos chamar de MRD (módulo de raciocínio dedutivo),

é especializado em inferência dedutiva e é anatomicamente separado de MIBC. Vários estudos têm encontrado evidência para a Teoria do Processo Dual do Raciocínio. Uma evidência desse tipo foi encontrada por Monti *et al* através de observações do cérebro por meio de imageamento por ressonância magnética funcional (MONTI, PARSONS & OSHERSON, 2009).

Os sujeitos no experimento de Monti e seus colaboradores tinham que avaliar raciocínios lógicos dedutivos e linguísticos enquanto tinham o cérebro imageado. O imageamento mostrou que o cérebro dos sujeitos tinha duas redes neurais distintas que entravam em ação durante os testes. Uma delas (que aparecia em verde nas imagens) era acionada quando os sujeitos avaliavam raciocínios dedutivos, e a outra (que aparecia em azul) era ativada quando os sujeitos julgavam raciocínios não dedutivos baseados em transformações sintáticas (especificamente, transformações da voz ativa para a voz passiva). Identificou-se ainda áreas ativadas em todos os testes, o que sugeria que elas eram áreas de apoio requeridas para a realização das duas tarefas cognitivas. A conclusão do estudo foi de que as áreas em verde correspondem à base neurofuncional de MRD e as áreas em azul correspondem à base neurofuncional de MIBC.

A ideia de que temos um módulo cognitivo inato especializado em raciocínio dedutivo levanta algumas questões interessantes. Vou encerrar esta seção discutindo três dessas questões, quais sejam: 1. Se já temos um MRD que nos permite fazer inferências dedutivas logicamente corretas, então para que serve a lógica?; 2. Se temos um MRD, como é possível que às vezes sejamos incapazes de distinguir deduções válidas de deduções inválidas?; 3. Como MRD funciona exatamente? As duas primeiras dessas questões têm respostas razoavelmente consensuais, enquanto que a última é objeto de uma acalorada controvérsia.

Em resposta à primeira questão, podemos dizer que a lógica é necessária porque ela funciona como uma teoria da inferência dedutiva e essa teorização da dedução nos permite uma expansão extraordinária do domínio de aplicações do raciocínio dedutivo. Explicando melhor, o que acontece é que nossa capacidade dedutiva inata nos fornece os padrões dedutivos elementares e isso nos permite fazer raciocínios dedutivos eficientes para as tarefas cognitivas da vida diária, mas a lógica nos permite fazer raciocínios mais complexos. O que a lógica nos dá é de certa forma semelhante ao que a aritmética nos dá. Sempre poderíamos fazer contas usando os dedos ou instrumentos como o ábaco, mas a aritmética nos permite edificar construções matemáticas mais sofisticadas. Da mesma forma, a lógica nos permite realizar operações lógicas muito mais sofisticadas do que aquelas que realizamos instintivamente, e graças ao simbolismo lógico, podemos inclusive programar máquinas para realizar tais operações.

À segunda questão, pode-se responder mostrando que há vários fatores que podem comprometer nossa capacidade de avaliar a validade das

deduções a despeito da acurácia do MRD. O que precisamos levar em conta é que:

a) O MRD dá o resultado certo quando recebe os dados certos, mas isso nem sempre acontece. Podemos errar ao representar a informação inicial e isso muito provavelmente vai provocar um erro no resultado final.
b) O MIBC concorre com o MRD. Podemos ser influenciados pela atmosfera dos dados iniciais e aplicar o MIBC em casos em que o correto seria aplicar o MRD. Como o MIBC não é apto para fazer ou avaliar inferências dedutivas, obtemos um resultado equivocado.
c) O funcionamento do MRD exige um uso razoável da memória de trabalho. Uma dedução mais complexa é geralmente feita em passos que precisam ficar retidos na memória. Desse modo, se faltar espaço na memória, a operação de dedução pode não ser bem sucedida.

Finalmente, em relação à terceira questão, é importante esclarecer que há duas teorias principais competindo para respondê-la. De acordo com a primeira delas, a Teoria dos Modelos Mentais, uma teoria que se iniciou com algumas pesquisas de Philip Johnson-Laird, o MRD é uma estrutura que manipula modelos mentais, sendo que, nessa concepção, um modelo mental é uma representação icônica (não proposicional) dos dados envolvidos na dedução. Diferentemente, para a segunda teoria, a teoria PSYCOP, de Lance Rips, o MRD manipula principalmente regras de dedução formais internalizadas que se caracterizam por sua simplicidade e automatismo. Houve um debate intenso entre Johnson-Laird e Rips, com ataques mútuos e defesas elaboradas, mas não vou aqui entrar nos detalhes do debate e das teorias quem estiver interessado em se inteirar melhor sobre assunto, deve ver (RIPS, 1994 e 1997, e JOHNSON-LAIRD, 1997a e 1997b). Aqui, basta-me indicar que a questão de como o MRD funciona exatamente ainda não está resolvida, e que as respostas mais cotadas são dadas por duas teorias cognitivas rivais. Isso é o bastante porque o que importa para meus propósitos é mostrar que a existência do MRD é hoje amplamente aceita na psicologia cognitiva, seja qual for o modo como se tenta explicar o seu funcionamento.

O que a Filosofia pode dizer disso tudo

Voltemos ao paradoxo de Carroll sobre Aquiles e a tartaruga e pensemos como as pesquisas da ciência cognitiva sobre a dedução poderiam nos ajudar a resolvê-lo? Não é difícil ver como. A hipótese de que temos mecanismos cognitivos designados especialmente para a realização e a avaliação de inferências dedutivas nos sugere a ideia de que tudo o que alguém precisa para se convencer da validade de uma inferência sancionada pela lógica é usar seu

módulo de raciocínio dedutivo. De acordo com essa ideia, portanto, para a tartaruga, a fonte da justificação de uma regra não pode vir de Aquiles, tem que vir de seus próprios mecanismos internos de dedução. Se a tartaruga estiver neurobiologicamente programada para reconhecer a regra, ou se a regra puder ser definida com base em regras mais simples que a tartaruga esteja programada para reconhecer, ela não poderá recusar sinceramente as inferências que as regras propostas por Aquiles autorizam.

Com isso, fica imediatamente claro como a explicação cognitiva da dedução nos possibilita dar uma resposta adequada à questão sobre a justificação dos preceitos da lógica dedutiva. Esses preceitos não são justificados pelo fim que eles nos permitem alcançar, quer pensemos que o fim seja a analiticidade ou a validade das deduções, pois ainda precisaríamos explicar o que nos leva a estabelecer tais fins, e tampouco são justificados pelo fato da lógica tentar sistematizar regras de inferência que já são usadas na prática dedutiva aceita pela comunidade de falantes, pois ainda precisaríamos explicar qual é o critério usado pelo lógico para incorporar no seu sistema algumas regras de inferência pragmaticamente sancionadas e excluir outras. Os preceitos lógicos são justificados pelo fato de que, em certo nível, nosso MRD nos faz *ver* que eles são corretos. Dito de modo direto, nós simplesmente somos programados para aceitá-los. Mas neste ponto surgem pelo menos duas dificuldades.

Em primeiro lugar, há o problema da pluralidade lógica. Não existe apenas uma única lógica dedutiva e, portanto, os axiomas e regras deferidos por uma lógica podem ser desautorizados por outra. Em vista desse relativismo lógico, parece inadequado sustentar que os preceitos lógicos estão fundados na nossa estrutura cognitiva. Pode-se argumentar que, se fosse assim, todos aceitaríamos as mesmas verdades lógicas e a lógica seria única.

Para encontrarmos uma solução para esse problema, creio que devemos pensar no porquê de existirem várias lógicas. Por que, por exemplo, a lógica intuicionista rejeita a regra da dupla negação? Claramente, isso ocorre porque a interpretação BHK da negação não nos permite concluir que uma fórmula α é verdadeira quando a fórmula$\neg\neg\alpha$ é verdadeira. Então, o que acontece é que, quando temos em mente a interpretação clássica da negação, reconhecemos a dupla negação como válida, e quando temos em mente a interpretação intuicionista, reconhecemos que a dupla negação é inválida. Temos a capacidade de reconhecer as duas coisas. E isso parece explicar a principal razão de existirem várias lógicas. Descobrimos lógicas desviantes quando assumimos semânticas desviantes.

O que deve ser notado, porém, é que os preceitos de uma lógica desviante ainda vão nos parecer aceitáveis em face da sua semântica. Isso parece indicar que a pluralidade da lógica é compatível com a explicação cognitiva. O que sabemos sobre MRD é que ele nos permite fazer e reconhecer deduções elementares, ou seja, ele nos permite receber certos dados de entrada e a

partir daí produzir certos dados de saída. Mas, se aceitarmos o paradigma representacional do processamento cognitivo, temos que entender que tanto os dados de entrada como os dados de saída são representações mentais. O que essas representações representam? Aparentemente, se nós podemos reconhecer tanto regras da lógica clássica como regras lógicas desviantes, nós representamos as proposições (quando raciocinamos a partir de proposições) sobre o fundo de uma semântica determinada. Consequentemente, podemos dizer que MRD nos possibilita ver a pertinência de uma inferência dedutiva sempre em relação ao contexto em que ela é feita.

Um exemplo pode nos ajudar a entender ainda melhor esse ponto. Louis Rougier uma vez considerou a afirmação: "todo número maior do que 99 é escrito com recurso a pelo menos 3 algarismos" (ROUGIER, 1941, p. 152). Essa afirmação é verdadeira? Se você usar o sistema decimal para escrever o número, a resposta é 'sim', mas se você usar o sistema hexadecimal a resposta é 'não'. Um sistema computacional simplesmente não pode dar uma resposta à pergunta se não considerar a informação a respeito do sistema numérico. Da mesma forma, MRD não pode dar uma resposta a respeito da validade da lei do terceiro excluído, por exemplo, se não considerar alguma informação sobre o modo de interpretar a disjunção e a negação. Agora, se ele dispor de toda a informação que precisa, ele não terá nenhum problema de fornecer uma resposta. Vemos, assim, que é perfeitamente possível defender que os princípios lógicos têm uma justificação cognitiva e ao mesmo tempo reconhecer que diferentes lógicas admitem e rejeitam diferentes princípios.

A segunda dificuldade concernente à explicação cognitiva da dedução pode ser expressa através de uma pergunta: nossos cérebros poderiam ser programados de outro modo? Podemos imaginar que eles são programados, por exemplo, para reconhecer o *modus ponens* como um tipo válido de inferência e, como só podemos raciocinar por meio dessa programação, não podemos ver como o *modus ponens* poderia ser inválido. No entanto, se tudo é uma questão de como a fiação do nosso cérebro está configurada, é possível imaginar que uma configuração diferente poderia nos compelir a rejeitar o *modus ponens* e a reconhecer uma regra muito diferente. Consideremos, por exemplo, a seguinte regra que chamo de *modus spurius*[1]:

$$\begin{array}{l} \alpha \\ \underline{\alpha \rightarrow \beta} \\ \neg\beta \end{array}$$

Seria possível que fôssemos programados para reconhecer a validade do *modus spurius*? (Note-se que o *modus spurius* não é preservador da verdade,

[1] A regra de inferência que chamo aqui de modus spurius é a mesma que Lance Rips chamou de modus shmonens no Prefácio de seu livro Psychology of Proof (cf. RIPS, 1994).

pelo menos não se atribuirmos à implicação e à negação os seus significados clássicos). Considerando o que foi dito anteriormente sobre a base neurobiológica do nosso instinto lógico, creio que a resposta mais coerente é 'sim'. Eu diria que essa é uma resposta que podemos conceber, embora não possamos imaginar como seria a sensação de acreditar no *modus spurius*. Em todo caso, a resposta é concebível porque é concebível uma situação hipotética em que, por alguma razão, um neurocientista do futuro de fato reprograma o cérebro de um sujeito para que ele aplique *modus spurius* em vez de *modus ponens*.

É importante reparar, porém, que essa é uma possibilidade cuja efetivação depende de um evento não natural. Se pensarmos apenas sobre o que seria possível em face da seleção natural que os seres vivos enfrentaram ao longo da história evolutiva, a resposta mais plausível à pergunta do parágrafo anterior parece ser 'não'. Temos boas razões para acreditar que nosso módulo de raciocínio dedutivo implementa as deduções que efetivamente precisamos fazer para sobreviver no planeta Terra. Com efeito, se nossa arquitetura cognitiva tem utilidade prática, então nossas regras de inferência internas devem ser tais que nos possibilitem resolver problemas práticos para nossa ação no mundo. Desse modo, um mutante que implementasse regras diferentes provavelmente não sobreviveria por muito tempo. Em particular, um mutante que implementasse o *modus spurius* seria extinto bem rapidamente. Isso fica mais claro se examinarmos um exemplo.

Digamos que você compre um bolo e acredite que se guardar um pouco, poderá comer um pedaço mais tarde, e que, além disso, você queira comer um pedaço mais tarde. O que você deve fazer? Parece óbvio que você deve guardar um pouco. Mas isso só é óbvio porque você é capaz de usar *modus ponens*. Se você usasse *modus spurius* e guardasse um pouco, a conclusão necessária é a de que você não poderia comer um pedaço mais tarde. Isso parece mostrar que, de um ponto de vista prático, o *modus ponens* é mais vantajoso que o *modus spurius* (também em tarefas mais abstratas, como, por exemplo, quando se preenche um quadro de sudoku, o aplicador de *modus ponens* leva vantagem). Um exemplo mais dramático pode mostrar que o *modus ponens* não só é mais vantajoso como é indispensável para a sobrevivência do indivíduo. Imagine, por exemplo, um pastor que raciocina que se cavar em certo lugar, achará água. Dado que ele precisa de água e que ele raciocina de acordo com o *modus ponens*, ele cava, e isso lhe possibilita matar sua sede. Já um pastor que raciocina de acordo com o *modus spurius* provavelmente morrerá de sede. E, de modo geral, sempre que fazemos predições condicionais e agimos com base nessas predições, fazemos isso porque usamos *modus ponens*. Se trocássemos o *modus ponens* pelo *modus spurius*, toda a nossa capacidade de fazer planos para o futuro e agir de acordo com esses planos ficaria irremediavelmente prejudicada. Uma explicação de por que isso acontece parece ser a de que os princípios de dedução realizados em nossa malha neuronal foram selecionados ao

longo de nossa história evolutiva. Em outras palavras, em última instância, nós raciocinamos como raciocinamos porque a realidade, pelo menos a realidade de nossa experiência empírica, exige isso de nós (se fôssemos seres subatômicos inteligentes, talvez raciocinássemos de forma diferente).

Essas conclusões parecem nos levar a especulações ainda mais ousadas. De fato, o que nosso módulo de raciocínio dedutivo faz é nos dotar com a capacidade de ver que nossa representação dos dados de entrada está indissociavelmente conectada a nossa representação da conclusão. Vemos isso em virtude de um tipo de reação biológica involuntária: quando o MRD recebe certa representação dos dados de entrada, ele nos compele para certa representação da conclusão. Assim, em última instância, quando dizemos, por exemplo, que β se segue necessariamente de α e de $\alpha \rightarrow \beta$, dizemos isso porque nos sentimos compelidos a representar β no momento em que representamos α e $\alpha \rightarrow \beta$. Destarte, nossa compulsão é o sinal indicativo que nos permite identificar a necessidade.

Isso significa então que podemos reduzir a necessidade lógica a um tipo de compulsão psicológica? Creio que essa não é a conclusão apropriada aqui. Se nossos módulos cognitivos estão ajustados às demandas da realidade externa e, em particular, nosso módulo de raciocínio dedutivo responde adequadamente a certas exigências práticas da vida, então o fato de representarmos certas inferências como inferências necessárias talvez deva ser interpretado como uma indicação de que há de fato relações necessárias na realidade, relações entre propriedades formais de estados de coisas. Em outras palavras, o que estou sugerindo é que talvez tenhamos sido programados para ver relações necessárias entre representações porque de fato existem relações necessárias entre os aspectos da realidade que representamos. Não digo que essa hipótese seja certa, mas ela certamente é plausível. Tão plausível quanto é inferir que há flores de corola tubular a partir da observação do bico de algumas espécies de beija-flores.

Termino assim com uma sugestão metafísica mais arrojada do que se poderia esperar de um artigo que trata do problema da justificação dos princípios da lógica a partir de certas evidências científicas para o nativismo lógico. Isso acontece porque vejo a explicação cognitiva em primeiro lugar como uma explicação da nossa capacidade de reconhecer a necessidade das inferências lógicas, mas também como uma ponte que pode nos levar a entender melhor a necessidade em si. A ideia, em suma, é esta: a aceitabilidade das regras lógicas se deve a algo que encontramos no nosso cérebro, mas o que encontramos no nosso cérebro se deve aparentemente a algo que encontramos no mundo.

Referências

BRAINE, M.D.S., O'BRIEN, D. P. (Eds.). *Mental logic*. Erlbaum, 1998.

CARROLL, L. What the Tortoise Said to Achilles. *Mind*, n. 4, 1895, p. 278–80.

CHATEAUBRIAND FILHO, O. *Logical Forms. part 1:* truth and description. Campinas: UNICAMP, 2001. (Coleção CLE, v.34).

CHOMSKY, N. "A Transformational Approach to Syntax". In: FODOR, J & KATZ, J. (Eds). *The Structure of Language: readings in the philosophy of language*. New Jersey: Prentice Hall, Inc., p. 211-245, 1964.

CHOMSKY, N. *Essays on Form and Interpretation (Studies in Linguistic Analysis)*. Elsevier Science Ltd, 1977.

CHOMSKY, N. *Language and the Problems of Knowledge:* The Managua Lectures. Cambridge, MA: MIT Press, 1988.

CHOMSKY, N. *New horizons in the study of language and mind.* Cambridge: Cambridge University Press, 2000.

CRAIN, S. *et al*. Quantification Without Qualification. *Language Acquisition*, v. 5(2), 1996, p. 83-153.

CRAIN, S. & KHLENTZOS, D. The Logic Instinct. *Mind & Language*, v. 25, n. 1, 2010, p. 30–65.

EVANS, J. St. B. T. In Two Minds: Dual-Process Accounts of Reasoning. *Trends in Cognitive Sciences*, v. 7, n. 10, 2003, p. 454-459.

GOODMAN, N. *Fact, Fiction and Forecast*. 4. ed. Harvard University Press, 1983.

JOHNSON-LAIRD, P. N. *Mental models:* Towards a cognitive science of language, inference, and consciousness. Cambridge, MA: Harvard University Press, 1983.

_____. Rules and Illusions: a critical study of rips's the psychology of proof. *Minds and Machines*, n. 7, 1997a, p. 387–407.

_____. An End to the Controversy? A Reply to Rips. *Minds and Machines*, n. 7, 1997b, p. 425–432.

JOHNSON-LAIRD, P. N. *How we reason*. Oxford, UK: Oxford University Press, 2006.

_____. & BYRNE, R. M. J. 'If' and the problems of conditional reasoning. *Trends in Cognitive Sciences*, v. 13, n. 7, 2009, p. 282–287.

MONTI, M. M., PARSONS, L. M. & OSHERSON, D. N. *Proceedings of the National Academy of Sciences of the United States of America (PNAS)*, v. 106, n. 30, 2009.

PRAWITZ, D. Remarks on some approaches to the concept of logical consequence. *Synthese*, v. 62, n. 2, 1985, p. 153-171.

RIPS, L. *The Psychology of Proof:* Deductive Reasoning in Human Thinking. A Bradford Book,1994.

RIPS, L. Goals for a theory of deduction: reply to Johnson-Laird. *Minds and Machines*, n. 7, 1997, p. 409–424.

ROUGIER, L. The Relativity of Logic. *Philosophy and Phenomenological Research*, v. 2, n. 2, 1941, p. 137-158.

Revista de Filosofia
ARGUMENTOS

Léo Peruzzo Júnior*

Linguagem e mente na Filosofia de Wittgenstein

RESUMO

Este artigo pretende analisar de que modo é possível falar, em Wittgenstein, da existência de um *estado interior* quando adotamos os recursos expressivos da linguagem. Neste sentido, os argumentos de Wittgenstein, especialmente em *Investigações Filosóficas* e *Últimos Escritos sobre a Filosofia da Psicologia*, permitem libertar a filosofia da mente de uma compreensão que insiste em separar o físico e o mental, enquanto distintos e independentes em substâncias e qualidades. Em linhas gerais, a primeira objeção à filosofia da mente consistiria em alegar que os modelos artificiais da cognição humana são capazes de replicar características específicas da vida mental humana como, por exemplo, é o caso das *qualia*. Uma segunda objeção, sustentada no decorrer no artigo é clarear, por um lado, a confusão gramatical e os pseudoproblemas que são associados à expressividade das *vivências interiores* e, por outro, estabelecer uma crítica ao modelo funcionalista de mente. Por fim, apontamos que a ambiguidade na expressão do conteúdo mental [ou significação do conteúdo mental] passa a residir nas sutilezas epistemológicas, e não ontológicas, da relação entre *linguagem*, *mente* e *sociedade*.

Palavras-chave: Linguagem; Mente; Sociedade; Filosofia da Mente; Wittgenstein.

ABSTRACT

This article aims to analyze how it is possible to approach, in Wittgenstein, the existence of an *inner state* when we adopt the expressive resources of language. In this sense, the arguments of Wittgenstein, particularly in *Philosophical Investigations* and *Last Writings on the Philosophy of Psychology*, enable the detachment of the philosophy of mind from an understanding that insists to

* Doutor em Filosofia pela Universidade Federal de Santa Catarina – UFSC. Professor Colaborador no Programa de Pós-Graduação Mestrado e Doutorado em Filosofia da Pontifícia Universidade Católica do Paraná – PUCPR. Professor da FAE Centro Universitário Franciscano do Paraná. Email: leo.junior@pucpr.br

separate the physical and the mental as distinct and independent in substances and qualities. In general, the first objection to the philosophy of mind would consist in claiming that the artificial models of human cognition are able to replicate specific characteristics of the human mental life such as the case of the *qualia*. A second objection supported throughout the article is, on the one hand, the clarification of the grammatical confusion and the pseudo-problems that are associated to the expressivity of the *inner experiences* and, on the other hand, to establish a criticism to the functionalist model of mind. Finally, we point out that the ambiguity in the expression of the mental content [or signification of the mental content] is in the epistemological niceties (not ontological) of the relation between *language*, *mind* and *society*.

Keywords: Language; Mind; Society; Philosophy of Mind; Wittgenstein.

Introdução

> O interior é uma ilusão. Isto é: o complexo de ideias aludido por essa palavra é como uma cortina pintada retirada da frente da cena do uso efetivo dessa palavra. (WITTGENSTEIN, *Últimos Escritos sobre a Filosofia da Psicologia*).

Quando G. Ryle criticou o dualismo cartesiano, especialmente na obra *The Concept of Mind*, sugerindo que o mesmo trata a "mente" como nome de um tipo de coisa específica, na verdade argumentava que, de fato, é apenas uma forma de se referir a certas *propriedades* e *relações* que seres humanos realizam habitualmente. De acordo com o argumento de Ryle, os enganos sobre as categorias transformaram-se em equívocos na tradição filosófica, uma vez que a apropriação destes conceitos, por exemplo o de "mente", permanecem num nível puramente abstrato e teórico (RYLE, 1951, p. 26-27). Ryle pretende, por sua vez, chamar a atenção sobre esses usos cotidianos dos "conceitos mentais" e, consequentemente, dos problemas que estes *usos* podem acarretar no materialismo moderno como afirma, por exemplo, em *Expressões Sistematicamente Enganadoras* (RYLE, 1975).

Se seguirmos os argumentos de Ryle e, deste modo, examinarmos a forma como o termo "mente" é utilizado, podemos evitar o que o autor denomina de *absurdidade lógica*, isto é, não haveria apenas dois tipos de coisas [material e mental] para compreendermos as diferentes descrições sobre o "mundo". Embora a posição de Ryle tenha sido geralmente descrita como *behaviorista* [um contra-movimento na psicologia insatisfeito com o método instrospeccionista] seus argumentos não pretendem atribuir apenas à *terminologia linguística* a resolução do problema. Seu objetivo é remover o *interior* de sua inacessibilidade [o aspecto

subjetivo da mente], demonstrando que não se trata apenas de um *erro linguístico*, mas, sobretudo de um *erro epistemológico* na sua descrição.

A cadeia de justificações traçadas por Ryle alerta para um problema fundamental em filosofia da mente: O que faz como que a *linguagem* seja significativa cognitivamente e, portanto, possa ser distinta de outros elementos [sons, movimentos, etc.] que ocorrem entre os seres humanos? Esta pergunta nutre, em si mesma, uma série de obstáculos sobre, por um lado, a natureza da relação entre a "linguagem" e a "mente" e, por outro, sobre o modo como temos acesso a ela. É neste cenário que Ryle, um herdeiro da filosofia ordinária de Wittgenstein ou, mais especificamente, de suas observações sobre *filosofia da psicologia*, situa os problemas que podemos considerar o cerne da dissolução ao problema *mente/corpo*[1], que abordamos no decorrer deste trabalho.

As explicações sobre a "mente" têm estabelecido definições que se iniciam com a filosofia e perpassam o campo das intituladas *ciências cognitivas,* que pretendem "desenvolver simulações de atividades mentais humanas", sendo "basicamente, uma ciência do artificial, ou seja, do comportamento das simulações entendidas como grandes experimentos mentais." (TEIXEIRA, 2004, p.13). Apresentada como *mito* [McGINN, 1991] ou, pelo contrário, como origem daquilo que realmente nos torna *humanos*, a "mente" tem despertado um interesse peculiar, especialmente nas últimas décadas, sobretudo por ter recebido abordagens que vão desde as preocupações *fisicalistas* [PLACE, 1956; SMART, 2004], perpassando pelo *materialismo eliminativista* [CHURCHLAND, 1979] ao *naturalismo biológico* [SEARLE, 2002], entre outras. Deste modo, as divergências entre as teorias podem ser tematizadas por duas grandes preocupações: a primeira, na necessidade de explicar como fazer uma tradução entre aquilo que ocorre em nosso *interior* e a sua relação cognitiva com o mundo exterior e, a segunda, em explicar a (im)possibilidade da existência de uma natureza mental ou, ao contrário, dar-lhe um caráter meramente físico-funcional.

Pretendemos mostrar, neste artigo, de que maneira é possível falar de uma possível visão *interior* num contexto de linguagem que assume características pragmáticas, o que implicaria, por exemplo, na eliminação da teoria funcionalista como modelo explicativo para os fenômenos mentais, uma vez que a atividade sintática não conseguiria aproximar-se da atividade humana consciente por não contemplar certos aspectos do discurso [semântica]. Para estabelecermos algumas hipóteses sobre "o que é a mente" e, consequentemente, "aquilo que ela não poderia não ser" resgatamos alguns argumentos da filosofia da psicologia de Wittgenstein, especialmente parte de seus escritos tardios em *Investigações Filosóficas* e *Últimos escritos sobre Filosofia da Psicologia*.

[1] É importante notar que Ryle não concentra seus argumentos em "mentes" e "corpos", mas em *seres humanos* como criaturas que pensam, sentem, etc. como qualquer outra atividade cotidiana, por exemplo, jogar uma partida de críquete. Neste sentido, a subjetividade não precisaria de uma dimensão interna para existir (RYLE, 1951).

Deste modo, o que haveria nos escritos de Wittgenstein, sobre a questão do *interior*, que servem como um possível diagnóstico para as teorias em filosofia da mente? É importante frisarmos que, especialmente nos escritos tardios de Wittgenstein, a *linguagem* é entendida como a chave da atividade que une o interno e o externo, sem que isso implique qualquer visão dualista sobre o tema. Esta condição aponta, de imediato, para um movimento que pode ser visualizado, por exemplo, nos recentes trabalhos de Searle (2002, p. 416), aqui especificamente *Consciousness and Language*, onde afirma que "a linguagem é realmente pública, e não depende do 'significado como uma entidade introspectível', dos 'objetos particulares', do 'acesso privilegiado' nem de nenhuma outra parafernália cartesiana". Sendo assim, nosso objetivo é mostrar que o descortinamento do *interior* pela linguagem, exposto nos escritos tardios de Wittgenstein, torna-se uma tentativa de torná-lo um estado não mítico, privado e fonte de ilusões [especialmente aquelas de natureza linguística].

Os fundamentos do materialismo moderno sobre o conteúdo mental

Em seus escritos sobre *filosofia da psicologia*, Wittgenstein parece claro, por um lado, não ter como objetivo discutir ou analisar os pressupostos epistemológicos utilizados pela psicologia de sua época, a saber, o possível *cientificismo da psicanálise* ou a *metodologia anti-introspeccionista do behaviorismo* de Watson. Por outro lado, Wittgenstein limita-se a uma interrogação gramatical, a uma investigação sobre o estatuto de certas palavras tais como *ver*, *sentir*, *desejar* que caracterizam os chamados "estados psicológicos". Vale notar que seu interesse está no problema da *significação*, que diz respeito a componentes externos e internos aos seres humanos. Segundo Gil de Pareja, a preocupação de Wittgenstein é descrever esta ligação "desde a análise dos termos até os enunciados que utilizamos para exteriorizar nossas vivências internas" (GIL DE PAREJA, 2002, p.16), não reduzindo a mente, portanto, a uma visão subjetivista ou materialista.

Especificamente em *Últimos escritos sobre a Filosofia da Psicologia*, manuscritos datados entre os anos 1945 a 1949, Wittgenstein analisa as questões do *interior* e sua *exteriorização* apontando uma abordagem que visa desconstruir uma leitura behaviorista sobre o tema em questão. Para demonstrar uma relativa desconfiança nos propósitos da referida teoria, o filósofo utiliza-se da "dissimulação" dos estados mentais para inferir que a análise apenas do comportamento externo pode ser inverossímil[2]. Deste modo, tentando evitar um embate filosófico entre *interior/exterior*, procura incidir, pelo menos em tese, para

[2] O conceito de "dissimulação" é utilizado, por Wittgenstein, numa série de exemplos ao longo de *Investigações*, especialmente a Segunda Parte (WITTGENSTEIN, 1996, Parte II).

algo que é exterior ao sujeito, mas que é condição necessária para sua determinação. Tal reflexão, se assim podemos nos referir, é que agora são reflexões sobre a atividade psicológica desde sua instância concreta, isto é, a *linguagem*.

As indagações feitas por Wittgenstein sobre a natureza, os fundamentos e o alcance da linguagem reconhecem o estado de confusão conceitual que afeta a utilização dos *conceitos psicológicos* e seu tratamento dentro da Psicologia. E esta questão parece ficar mais evidente quando Wittgenstein aborda o estatuto dos verbos psicológicos como, por exemplo, *crer, desejar, esperar*, para mostrar que há por detrás uma natureza linguística que deve ser desvelada. Com isso, se retomamos a posição de Ryle, sob a hipótese wittgensteiniana, as operações que a mente humana pode executar não podem ser apreendidas a partir de uma avaliação da própria consciência. E isso, portanto, ocorreria por duas questões: a primeira, porque a linguagem não é um movimento privado ou solipsista; a segunda, porque não seríamos um "fantasma na máquina".

A imagem de que o acesso ao *interior*, por um lado, esteja envolto por uma máscara e, por outro, tenha uma relação simétrica com o comportamento, simplesmente retira a "mente" de seu uso originário. A observação permite apontar que os conceitos psicológicos, utilizados para a descrição do conteúdo mental, não podem ser derivados de um universo extra ou meta social. Neste sentido, é necessário saber o que se fala ao utilizarmos palavras como "pensar", "perceber", "imaginar", "sentir", entre outras. (WITTGENSTEIN, 1994, p.19-21), uma vez que elas não são categorias, presumivelmente, arbitrárias à linguagem. Por exemplo, quando alguém parece esconder seus pensamentos tem-se a impressão de que o *interior* está oculto atrás de algo. Isso significa, erroneamente, segundo Wittgenstein, que haveria um processo misterioso que envolve o *interior* e estaria associada a sua ocultação como algo que se encontra fora da linguagem, além dos limites do mundo e implausível de cognição absoluta.

A vacuidade do termo "mente", em diversas situações ["Deixa ver se *consigo* lembrar!", "Eu fiz isso sem *pensar*", "Não era isso que eu queria dizer", etc.], e sua associação com super-conceitos ou falsas imagens ["a mente", "a consciência", etc.], acaba coincidindo com a tradução metafísica da existência de algo para além da linguagem. Assim, ao contrário de uma arbitrariedade da linguagem, como aponta Wittgenstein, suas regras não podem designar nenhuma coisa que esteja fora dela. A celeuma entre *objetividade* e *subjetividade*, entre *mente* e *corpo*, portanto, poderia ser explicada somente a partir de uma digressão histórica, retratada pela crença de que, em última instância, a *ciência* é exclusivamente uma propriedade empírica, eliminando os paradoxos anteriores. Como consequência, em linhas gerais, na visão materialista moderna sobre a relação *mente/corpo*, residiriam alguns argumentos fundamentais:

- **Argumento 1.** Os termos mentais exprimem disposições comportamentais, onde os termos mentalistas são sinônimos dos termos disposicionais;
- **Argumento 2.** As causas mentais ocasionam efeitos comportamentais em virtude de outras causas mentais;
- **Argumento 3.** Os eventos e estados mentais são idênticos a processos neurofisiológicos do cérebro, ou seja, a propriedade de certo estado mental é idêntica a certo estado neurofisiológico.

Em que sentido seria possível aceitarmos o materialismo e, deste modo, que as características especiais do significado humano sejam derivadas do nosso uso da linguagem? Na filosofia da psicologia, a atenção de Wittgenstein gira em torno da linguagem, uma vez que os conceitos relativos às experiências interiores conectam-se diretamente com a atividade humana, mostrando que o *interno* é produto de tal relação. Neste caso, o *interior* pode ser melhor compreendido quando se desfazem as ficções gramaticais originadas nos "conceitos de direto e indireto, tais como a de que temos acesso direto a nossas dores ou de que temos apenas acesso indireto à dor de um outro, enquanto ele tem acesso direto à sua própria dor" (HEBECHE, 2002, p. 85). Embora os fenômenos do mundo da consciência, como geralmente se acredita, são *subjetivos* e *privados*, isso não significa afirmar que eles possam ser algo excepcional diante da matéria que compõe o mundo. Assim, parece que uma taxonomia das operações mentais reforça o argumento de que há muitos elementos imbricados entre "a mente" e onde ela, de fato, deve ocorrer:

> 1) As percepções externas das coisas e que identificamos e que nos cercam constantemente, e também as percepções "internas" (às vezes dizemos que "percebemos" coisas na imaginação, na memória e nos sonhos, que percebemos uma distinção, etc.);
>
> 2) As sensações de cores, texturas, timbres, etc., e as sensações que acompanham cada um de nossos movimentos e que chamamos de "propriocepções"; as dores e prazeres de várias intensidades que, infelizmente ou por nosso bem, sentimos constantemente. Temos aqui o domínio dos *qualia*, características qualitativas das experiências conscientes, presentes nas percepções;
>
> 3) As imagens mentais que acompanham atividades (mentais) como imaginar algo (existente ou inexistente), se lembrar, antecipar, etc.;
>
> 4) Atitudes proposicionais ou estados providos de conteúdo conceitual que podemos ter pontualmente ou durante certo tempo a título de disposição, como acreditar que a Seleção brasileira ganhou a Copa do Mundo de 2002, ter a intenção de viajar à China daqui a dois anos, desejar casar com a rainha de Tebas, etc.;
>
> 5) O domínio das emoções: sentir medo, recear, criar coragem, ficar triste ou alegre, se emocionar, sentir vergonha ou orgulho;
>
> 6) Atos ou operações como conceber, julgar, decidir, deliberar, raciocinar, ordenar, se lembrar, etc.;

> 7) As disposições, em geral, além das atitudes proposicionais já mencionadas: capacidades (como reconhecer os rostos), habilidades (falar uma língua, dirigir um carro, adicionar, dividir, multiplicar mentalmente, etc.), ou ainda ter senso de humor, ser honesto ou mentiroso, ser fumante, gostar da música de Handel, saber tocar piano, etc (LECLERC, 2010, p.16-17).

Mas para que, efetivamente, serve este quadro de descrições sobre os "estados mentais"? A resposta consiste em duas direções que não podem ser auto-eliminativas: a primeira, para mostrar que alguns "eventos mentais" dependem de uma covariação causal do funcionamento de um sistema biológico (exemplo, Argumento 1), enquanto outros envolvem a aplicação de conceitos e têm uma dimensão normativa (LECLERC, 2010, p. 17). De qualquer modo, inevitavelmente, posições diversas em filosofia da mente têm se apropriado, talvez de maneira pouco sensata, dos equívocos que assombram tais descrições.

O que vale para a argumentação anterior, é o fato de que Wittgenstein não realiza uma investigação sobre a *natureza do interior*, mas sobre o modo como efetivamos sua exteriorização por meio de uma linguagem, de caráter público, e pelo seguimento de regras. Assim, parece claro, especialmente nos primeiros aforismos dos *Últimos Escritos,* que os problemas conceituais a respeito do interior são criados a partir das armadilhas da linguagem na sua exteriorização. No caso da dor, por exemplo, Wittgenstein afirma que se não existissem critérios públicos, nunca compreenderíamos o que significa quando outra pessoa afirmasse ter dores (WITTGENSTEIN, 2007). Portanto, pode-se apontar que, segundo Wittgenstein, o interior não deve ser visto como uma *caixa preta* (*black box*), onde cada indivíduo parece esconder algo sobre suas vivências interiores. Ao contrário, o conteúdo mental herda propriedades semânticas e pragmáticas da linguagem que é utilizada para a instanciação da consciência.

Por que a posição de Wittgenstein sobre a "mente" é anti-behaviorista?

Divergente aos mentalistas [introspeccionistas] e dualistas, que supõem a existência de estados internos e representações que influenciam a determinação do comportamento, John Broadus Watson (1878-1958), considerado o fundador do *behaviorismo metodológico*, abandona o estudo dos processos mentais (por exemplo, *pensamentos* e *sentimentos*) e passa a descrever e analisar o processo psicológico por meio do comportamento exterior. Watson acreditava que era por meio deles que o homem se constituiria e, por esta razão, seria possível estabelecer a descrição e compreensão da "consciência".

O intento da psicologia, segundo Watson (1961), seria o de prever e controlar seu objeto de estudo, ou seja, o comportamento. Neste sentido, Watson compreende que a ideia de existência de uma vida mental é *superstição*, *um*

resquício da Idade Média. O behaviorismo em questão[3] tentava demonstrar que todos os fenômenos e eventos psicológicos, como nos exemplos descritos por Leclerc (2010, p.16-17), só podem ser analisados pela observação e previsibilidade dada pelo comportamento.

Deste modo, para sanar a digressão histórica, especialmente aquela cartesiana, Watson propõe fazer do *corpo* o objeto de estudo da psicologia, cabendo ao cientista opor-se às explicações de origem interna, ou mental, do comportamento humano. Este argumento permitiria retirar de cena a colaboração subjetiva e as explicações de cunho religioso ou metafísico. As primeiras propostas de Watson foram apresentadas em 1913, em um texto publicado na *Psychological Review*, intitulado *Psychology as the behaviorist views it*. Especialmente neste texto, o behaviorismo de Watson não negava a existência da *mente*, ou de algo *interior*, mas recusava seu estudo em razão de sua inacessibilidade e ausência de estatuto científico (WATSON, 1961, p. 158-177).

Na visão de Watson, a psicologia de Wundt (1896) apresentava-se, ainda, como um momento de transição entre o dualismo filosófico e a psicologia científica, o que não apontaria uma solução clara para o problema *mente/corpo*. Assim, por um lado, o behaviorismo passava a questionar a objetividade da consciência pela introspecção e, por outro, retirava do vocabulário da psicologia os termos subjetivos que estão além daquilo que se possa descrever na relação entre estímulo e resposta [E-R]. Assim, argumenta Watson: "Por que não fazer do que podemos observar o verdadeiro campo da psicologia? Limitar-se a observar e formular leis relativas somente a essas coisas. E que coisas podemos observar? Somente o comportamento – o que o organismo faz ou diz" (WATSON, 1961, p. 158). Os elementos que constituem o conteúdo mental [crenças, desejos, imagens mentais, etc.], por exemplo, não estariam sujeitos à experimentação científica.

Fica evidente que se, numa visão introspeccionista, na análise do comportamento verbal, há um experimentador e um observador que descreve suas experiências, para o behaviorismo de Watson (1961), o experimentador é o observador que relata suas experiências internas por meio da substituição dos objetos por palavras. Isso implica que o *comportamento observável* explica o comportamento (in)consciente de alguém, tornando a subjetividade uma propriedade descritível objetivamente. Especificamente sobre este argumento, é significativo frisar que os enunciados acerca de *estados mentais*, segundo Watson (1961, p.160-161), poderiam ser descritos na observação do comportamento, o que excluiria a possibilidade, por exemplo, de que uma manifestação

[3] A corrente behaviorista poderá também se apresentar em outras duas versões: o *behaviorismo metafísico*, que nega a existência de fenômenos mentais; e o behaviorismo lógico, que afirma que as proposições acerca do nível mental são semanticamente equivalentes a proposições acerca de disposições comportamentais.

de *tristeza*, ou *dor*, ou *crença*, etc. ser interpretada de maneira diferente em outra ocasião.

O mecanismo estímulo/resposta (E.....R) do behaviorismo de Watson confronta-se com os escritos de Wittgenstein, uma vez que a verificação do significado de uma proposição, apenas pelo seu comportamento externo, por exemplo, está distante de ser a única forma de compreendermos as condições de significação. Neste sentido, as observações de Wittgenstein sobre filosofia da psicologia não logram a cientificidade ou a materialidade dos eventos mentais. Ao contrário, seu interesse é uma interrogação de natureza gramatical, o que implica combater os reducionismos materialistas em filosofia da mente como, por exemplo, a corrente interessada em eliminar a *folk psychology* e instaurar uma "ditadura dos conceitos neurofisiológicos" (SMART, 2004, p.116). Numa das passagens de *Investigações*, Wittgenstein destaca:

> Não será você um behaviorista disfarçado? Você não diz que, no fundo, tudo é ficção, salvo o comportamento humano? – Se falo de uma ficção, trata-se então de uma ficção gramatical. (WITTGENSTEIN, 1996, §307).

Wittgenstein, portanto, manteria uma visão anti-behaviorista por deslocar o problema do comportamento para a questão da linguagem. A dinâmica expressiva desta, que o autor faz frente tanto ao modelo dualista apresentado quanto ao behaviorismo metodológico, são traços sinuosos que estão lado a lado na mesma moeda. Assim, como interpreta Putnam (2002, p.86) a respeito da posição wittgensteiniana,

> A rejeição do 'cartesianisno' cum 'materialismo' não significa [...] voltar ao próprio dualismo cartesiano. [...] O discurso mental se compreende melhor como discurso de determinadas aptidões que possuímos, aptidões essas que dependem do cérebro e de todas as inúmeras transações entre o meio ambiente e o organismo [...].

Por fim, o desfecho do behaviorismo pode ser expresso pela própria conclusão apresentada por Searle, em *Consciousness and Language*, ao afirmar que durante a fase positivista e verificacionista da filosofia analítica,

> não era difícil divisar a razão do desejo de eliminar o mental: se o significado de uma afirmação é o seu método de verificação, e se o único método de verificação das afirmações sobre o mental reside na observação do comportamento [...] então, "as afirmações sobre o mental são equivalentes, quanto ao significado, a afirmações sobre o comportamento." (SEARLE, 2002, p.336).

Assim, uma objeção importante ao behaviorismo é que o *conteúdo mental*, por um lado, não pode ser reduzido às regras sintáticas ou, por outro,

a uma linguagem puramente privada, no sentido de que apenas uma pessoa poder, em princípio, compreender.

Serão os estados psicológicos *Estados Internos*?

Pode parecer que a própria resposta deveria assemelhar-se, de algum modo, ao postulado de que os *estados internos* são *estados psicológicos*. Esta é a concepção funcionalista de um estado psicológico, isto é, um estado psicológico é um "estado funcional" que liga os estímulos a respostas sensoriais (PUTNAM, 2002, p. 190). De acordo com o funcionalismo, a natureza essencial dos eventos mentais [dores, desejos, crenças, etc.] não deve ser buscada na matéria de que são compostos, mas na *função* que cada um executa. Neste sentido, para reforçar as antinomias realistas, parece significativo que o funcionalismo de Putnam, seguindo os traços wittgenstenianos, pergunte-se "como é possível que a linguagem se encaixe no mundo?", ou ainda, "como a percepção se encaixa no mundo?" (PUTNAM, 2002, p. 35). A resposta estaria na "semântica verificacionista", isto é, na crença de que nossa linguagem deve consistir no domínio de uso da mesma, de onde situaríamos os estados psicológicos e as vivências interiores.

Mas, a que tipo de legado wittgensteiniano se deve a interpretação funcionalista de Putnam a respeito do *conteúdo mental*? Wittgenstein não se dirige ao estudo dos enunciados empíricos da Psicologia, "mas sua indagação se concentra em uma consideração gramatical dos usos dos termos e enunciados psicológicos tal como se encontram no seu uso ordinário" (GIL DE PAREJA, 1992, p. 75). Trata, por um lado, como ressalta Gil de Pareja, do problema da *linguagem privada*[4] e, por outro, das abordagens externalistas sobre o conteúdo mental. Estes dois pontos, portanto, incorporam uma tendência que supõe que o vocabulário com o qual expressamos nossos conceitos mentais (dor, ódio, amor, etc.) adquire significado em virtude da relação com nossas próprias experiências cotidianas, como concorda Putnam (2002).

Entretanto, caso levemos a sério a ideia anterior, deveríamos supor que o conhecimento do mundo interno, por parecer ser inacessível ao mundo externo, indicaria apenas o acesso exclusivo do próprio sujeito. Isto mostra que, de forma bastante genérica, teríamos *certeza* apenas do conhecimento de nosso *interior*, mas nunca poderíamos estar seguros dos pensamentos e sentimentos daquilo que são as vivências interiores de outras pessoas. Sendo assim, se o externalismo semântico estiver correto, portanto, estaríamos inclinados em adotar duas hipóteses: 1. que realmente não podemos saber o que acontece em outras *mentes*, uma vez que o significado de um termo estaria determinado por um estado psicológico particular; 2. que o significado está

[4] Cf. WITTGENSTEIN, 1996, §243-315.

envolvo por um *fenômeno social* e, portanto, seria sempre determinado por condições ambientais [por exemplo, a visão relativista de Richard Rorty (1979)].

Villanueva afirma que a resposta de Wittgenstein é que não podemos atribuir determinados signos externos como consequência causal de todo e qualquer estado mental (VILLANUEVA, 1996, p. 24). Neste caso, por exemplo, a falta de critérios públicos mostraria a possibilidade de uma interpretação errônea do estado mental. Saber que alguém está em um estado mental particular não é somente ter a capacidade de predizer como irá comportar-se na continuação do ato, mas, sobretudo, sermos capazes de entendê-lo. (VILLANUEVA, 1996, p. 25). Com isso, o *interior* não é um conjunto de objetos privados ou escondidos, afirma Wittgenstein. A indefinição dos conceitos psicológicos (por exemplo, crer, desejar, etc.), ou seja, a sua flexibilidade é "[...] a forma que permite a compreensão do interior por meio de sua expressão nos jogos de linguagem." (WITTGENSTEIN, 2007, p. 40); este elemento implica, portanto, a aproximação entre as *Investigações Filosóficas* e as notas que compõem os *Últimos Escritos sobre a Filosofia da Psicologia*. Se todo conceito psicológico tivesse como consequência a sua correta compreensão, então conseguiríamos decifrar e replicar todos os estados mentais e saberíamos a rigor o que acontece em outras mentes. Este argumento, sem sombra de dúvidas, tornaria falsa a crítica de Searle ao projeto da inteligência artificial forte (SEARLE, 2002, p. 110-111).

Por fim, é importante notar que, se nossas vivências internas [conteúdo mental] fossem observáveis apenas pelo comportamento externo não verbal, então, elas seriam semelhantes às vivências dos outros e, todo fenômeno privado se comportaria através da simples observação externa semelhante [e o behaviorismo lógico ou metodológico estariam corretos]. Em contrapartida, a insegurança em afirmar que compreendemos os estados privados de outras mentes possui equivalência em afirmar que não podemos ter certeza do que as outras mentes possam conhecer a nossa. Se corretas as hipóteses anteriores, os pilares da *certeza* e da *dúvida* sobre os processos cognitivos tornam-se cada vez mais instáveis. Por isso, como não temos razões suficientes para poder afirmar a existência de outras mentes, também não poderíamos negá-las ou, epistemologicamente falando, reduzir a *mente* apenas a descrição possível dada pela linguagem.

"Comportar-se" ou "Dissimular": a *Certeza* sobre outras *Mentes*

Segundo Wittgenstein, em sua *Filosofia da Psicologia,* a noção de *experiências privadas* será uma ilusão, uma espécie de miragem que coloca algo no interior do sujeito para lá da forma linguística. Não podemos inferir, conforme expõe Wittgenstein, algo com uma intencionalidade tal que entende o *interior*

como um ponto localizado e plausível de privacidade: "Evidentemente, existe um fragmento do jogo de linguagem que sugere a ideia de ser privado – ou de estar escondido – e existe também algo que pode denominar-se esconder o interno." (WITTGENSTEIN, 1996, p. 50). Nos apontamentos MS 169, escritos por volta de 1949, por exemplo, Wittgenstein pede que imaginemos como se estivéssemos em uma espécie de concha de caracol, e quando a nossa cabeça está para fora então nosso pensamento não seria privado, apenas quando a recolhemos (Cf. WITTGENSTEIN, 1996, p. 47). O objetivo deste exemplo é criticar a falsa impressão de que o movimento *interno/externo*, ou o contrário, elimine a capacidade de dissimular o comportamento. Isso significa que a possibilidade de expressão falsa de um *conteúdo mental* seriam um indicativo de que tanto o behaviorismo quanto o funcionalismo seriam insuficientes para a explicação da "mente": o primeiro, porque o comportamento pode ser dissimulado; o segundo, porque o *outputs* pode ser diferente do *inputs*.

Wittgenstein procura mostrar que é na concretude da gramática, na *força ilocucionária*, que se dá compreensão do conteúdo que compõe a "mente". A componente subjetiva é necessária para que alguém possa afirmar que sabe alguma coisa, mas não é suficiente, já que tem que indicar *razões* ou *justificações*, que são públicas [sociais] e sem as quais a sua convicção não deve ser considerada. Segundo Marques (2003, p. 136), Wittgenstein desenvolve uma visão panorâmica das gramáticas dos verbos cognitivos, defendendo uma noção consensualista de verdade. Isso contradiz o argumento de que, se os termos mentais adquirem significado a partir do próprio eu, então nossos conceitos de *tristeza* ou *dor*, por exemplo, são irredutivelmente subjetivos e seriam essencialmente privados no sentido que somente o sujeito que experimenta a *dor* ou *tristeza* pode saber se seu estado mental é correspondente ao ambiente externo.

É notável que, ao que parece, só podemos alcançar segurança cognitiva [uma espécie de compreensão definitiva], quando nos referimos ao nosso próprio interior, porque ao sentir *dor*, por exemplo, o único critério é a auto-observação, condição que nos levaria novamente a defender o introspeccionismo. Wittgenstein argumenta, nos *Últimos Escritos sobre a Filosofia da Psicologia*, que a correlação entre *interno* e *externo* não é, por sua vez, suficiente para explicar a existência do primeiro (interno): "Estou seguro que ele tem dores. O que significa isto? Como se usa? Qual a expressão de segurança na conduta que nos fazem estar seguros?" (WITTGENSTEIN, 1996, p. 32). Então, segundo o próprio autor, a evidência disponível a favor de um interior, de um estado mental, ou supostamente de uma "mente", seria a capacidade de *dissimular* um evento, quando na verdade não se o tem.

Esta capacidade de dissimular as experiências internas é admitida quando a conduta externa é fictícia, por exemplo, o fato de não ter dor e poder simular tal estado. Sendo assim, por um lado, se a relação entre o *mundo ex-*

terno e o *mundo interno* fosse causal, poderíamos replicar e prever o que são as outras mentes, uma vez que isso nos aproximaria da premissa funcionalista da Inteligência Artificial: "Os estados mentais são estados funcionais". Já por outro lado, num certo sentido, máquinas podem *pensar*, porém não poderiam *dissimular* como os seres humanos:

> O que explica mais bem a nossa insegurança na hora de atribuir estados mentais aos demais é o próprio jogo de linguagem da autodescrição. Não é que nossa insegurança se explique por uma espécie de vazio entre o interno, que está oculto, e o externo, a conduta, que é pública. O que acontece é que os critérios de conduta para essas autodescrições são constitutivamente indeterminados: nunca sabemos quando a evidência é suficiente para dizer que esta dor de dente é autêntica ou simulada, porque tais critérios carecem de limites definidos. Em consequência, a insegurança não pode eliminar-se porque, como afirmou-se, é parte do jogo de linguagem. (VILLANUEVA, 1996, p. 16).

A afirmação de Villanueva sobre Wittgenstein mostra que não poderíamos saber ou afirmar, *ipso facto*, o que acontece no *interior* das outras pessoas, ou seja, se realmente tal evento corresponde à vivência interna exteriorizada ou se ela está sendo dissimulada. Neste sentido, Wittgenstein descreve que o gênero da *certeza* depende do gênero do *jogo de linguagem* em questão: "Não pense em estar seguro com um estado mental, um gênero de sentimento, ou algo do estilo. O importante na segurança é a maneira correta de atuar, não a expressão da voz com que se fala." (WITTGENSTEIN, 1996, p. 32). Ao considerar a linguagem descritiva do conteúdo mental, ou das vivências privadas, como uma espécie de *jogo*, adverte que a compreensão de algumas palavras inclui a possibilidade de usá-las em certas ocasiões associadas a gestos ou com um tom especial de voz.

A simetria apontada por Wittgenstein, entre a primeira e terceira pessoas da linguagem, por um lado, rechaça a acusação behaviorista e, por outro, mostra que um enunciado da primeira pessoa preserva seu sentido quando é substituído por aquele da terceira pessoa. Não se está falando da simetria ou compatibilidade entre *mentes*, mas na ocorrência de cognição na forma como ocorre a descrição entre estas *mentes*. Sendo assim, a descrição das vivências internas tornar-se-á possível quando a linguagem descrever os conceitos do mundo interior de forma pública, ordinária (HEBECHE, 2002, p. 75), através da nossa *folk psychology* [e contrário ao materialismo eliminativista].

Considerações finais

Nos escritos de Wittgenstein em questão, a autonomia do *uso* determina a utilidade correta da linguagem que descreve os processos cognitivos, isto é, o autor identifica a *compreensão* não como um estado interno de conheci-

mento, mas como uma *ação comunicativa intersubjetiva*. O caráter peculiar das observações sobre *filosofia da psicologia*, em Wittgenstein, permitem sustentar uma crítica aos problemas epistemológicos com que o *mental* é geralmente tratado, especialmente pelas correntes fisicalistas, ou materialistas, nas neurociências, e behaviorista, na psicologia. Consequentemente, o que tudo isso revela é que parece ainda haver, na visão tradicional da filosofia da mente, um nível ortodoxo que continuará sustentando que os estados mentais distinguem-se dos demais por possuírem um conteúdo qualitativo.

A leitura de Wittgenstein permite apontar, contudo, duas elucidações aos dilemas tradicionais em questão: o primeiro, realiza um *diagnóstico* sobre o estado de confusão decorrente do mau uso da linguagem, especialmente quando falamos da possibilidade de expressão do conteúdo mental [crítica à *linguagem privada*]; e, segundo, refuta uma tendência muito geral e sempre presente que insiste numa suposta ontologia do *interior* como realidade distinta de toda experiência exterior humana. Esta última posição seria resguardar a *mente* a uma proposta metafísica que continua se arrastando nas ciências em geral, sendo alimentada pela tradição filosófica [por exemplo, a teoria dos aspectos qualitativos dos estados mentais sustentada por Nagel (1995)]. O resultado disso é que não podemos, em última análise, separar a sensação de dor da possibilidade de expressá-la de alguma forma acessível. Obviamente, ninguém poderia sentir dor, por exemplo, se não tivesse sensações de alguma espécie; mas, ninguém poderia sentir dor a menos que pudessem expressá-la publicamente. Que seja ou não possível resolver o paradoxo sobre a natureza do mental, nenhuma dessas possibilidades anteriores implica, analiticamente, a defesa de alguma espécie de dicotomia ou desconexão entre "mente" e "linguagem".

Referências bibliográficas

CHURCHLAND, Paul. *Scientific Realism and the Plasticity of Mind*.Cambridge University Press, 1979.

GIL DE PAREJA, José Luis. *La Filosofía de la Psicología de Ludwig Wittgenstein*. Barcelona: PPU, 1992.

HEBECHE, Luiz. *O mundo da consciência. Ensaio a partir da filosofia da psicologia de L. Wittgenstein*. Porto Alegre: EDIPUCRS, 2002.

LECLERC, André. Mente e "Mente". *Revista de Filosofia Aurora*, v. 22, n. 30, p. 13-26, jan./jun. 2010.

MARQUES, Antonio Carlos. *O interior:* linguagem e mente em Wittgenstein. Lisboa: Fundação Calouste Gulbenkian, 2003.

NAGEL, Thomas. *Other minds:* critical essays (1969-1994). Oxford: Oxford University Press, 1995.

McGINN, Colin. *The problem of consciouness.* Oxford: Blackwell, 1991.

PERUZZO JÚNIOR, Léo. *Wittgenstein:* o interior numa concepção pragmática. Curitiba: CRV, 2011.

PLACE, Ullin T. 'Is Consciousness a Brain Process?'. *British Journal of Psychology*, v. 47, p. 44–50, 1956.

PUTNAM, Hilary. *A tripla corda:* mente, corpo e mundo. Lisboa: Instituto Piaget, 2002.

RORTY, Richard. *The Philosophy and the mirror of nature*. Princeton: Princeton University Press, 1979.

RYLE, Gilbert. *The Concept of Mind*. Londres: Hutchinsons University Library, 1951.

_____. *Expressões Sistematicamente Enganadoras*. *Ensaios.* Tradução Balthazar Barbosa Filho. São Paulo: Abril Cultural, 1975. (Coleção Os Pensadores).

SEARLE, John. *Consciousness and language*. London: Cambridge University Press, 2002.

_____. PERUZZO JÚNIOR, Léo. *Mind, language and society in philosophy of John Searle* (Interview). Principia, v. 21, n.1, jan./abr. 2015.

SMART, J. J. C. Sensations and brain-process. In: HEIL, J. (Ed.). *Philosophy of mind*: a guide and anthology. Oxford: Oxford University Press, 2004, p. 116-127.

TEIXEIRA, João de Fernandes. *Filosofia e Ciência Cognitiva*. Petrópolis: Vozes, 2004.

VILLANUEVA, Luis Manuel Valdés. Estudio Prelininar. In: WITTGENSTEIN, Ludwig. *Últimos escritos sobre Filosofía de la Psicología. Lo interno y lo externo*. v. 2. Tradução Luis Manuel Valdés Villanueva. Madrid: Editorial Tecnos, 1996.

WATSON, John Broadus. *El conductismo*. Buenos Aires: Editorial Paidos, 1961.

WITTGENSTEIN, Ludwig. *Investigações filosóficas*. Petrópolis: Vozes, 1996.

_____. *Últimos escritos sobre a Filosofia da psicologia*. Tradução António Marques, Nuno Venturinha, João Tiago Proença. Lisboa: Fundação Calouste Gulbenkian, 2007.

WUNDT, Willem. Uber die definition der psychologie. *Philosophische Studien*, v. 12, p. 307-408, 1896.

ARGUMENTOS
Revista de Filosofia

Sascha Rammler*

What does extensionality show in the *Tractatus?*

ABSTRACT

Extensionality and extensionalism are common themes in Analytic Philosophy. The early Wittgenstein of the *Tractatus* is also taken to hold a thesis of extensionality. Extensionality in the *Tractatus* is associated with sentence 5, where Wittgenstein claims that a proposition is a truth-function of elementary propositions. The notion of a truth-function in the *Tractatus* is approached by an *operational view* in this paper that takes the truth-functions themselves as generated by a truth-operation. In this sense, the truth-operation is generating the notation itself, not an interpretation of some formal language. Extensionality in the *Tractatus* is approached in three steps, illustrating first what an operational reconstruction can show about extensionality, continuing with the role the *Tractatus* assigns to extensionality, and concluding by comparing it to other uses of the term extensionality in the Analytic tradition.

Keywords: Extensionality; Truth functionality; Early Wittgenstein; *Tractatus.*

RESUMO

Extensionalidade e extensionalismo são temas comuns em Filosofia Analítica. O primeiro Wittgenstein, o do *Tractatus*, também é tomado como defendendo extensionalidade. Extensionalidade no *Tractatus* é associada à sentença 5, onde Wittgenstein reivindica que a proposição é uma função de verdade de proposições elementares. Neste sentido, a operação de verdade gera a notação ela mesma, e não uma interpretação para alguma linguagem formal. Extensionalidade no *Tractatus* é abordada em três passos, a saber, ilustrando, primeiramente, o que uma reconstrução operacional pode mostrar sobre extensionalidade. Em seguida, analisando o papel que o *Tractatus* atribui à extensionalidade, e concluindo ao comparar esta propriedade com outros usos do termo extensionalidade na tradição analítica.

Palavras-chave: Extensionalidade; Vero-funcionalidade; primeiro Wittgenstein; *Tractatus.*

* PhD Student of the International Graduate School (IGS) at BTU-Cottbus-Senftenberg (Brandenburg University of Technology Cottbus-Senftenberg). sascha.rammler@googlemail.com

Introduction

The aim of this paper is to discuss the theme of extensionality in Wittgenstein's *Tractatus Logico-Philosophicus*[1]. I will focus mainly on sentence 5 and its sub-sentences or commentaries according to the *Tractatus* numbering system. The goal is to reconstruct the particular variety of extensionality presented in the Tractatus by the early Wittgenstein and to contrast it to some extent with other notions of extensionality.

Sentence 5 of the *Tractatus* is commonly referred to as a *thesis of extensionality* (Carnap 1937, p. 188, Black, 1964) p. 219, Frascolla (2007) p. 118. Rosenberg (1968) p. 341). The *Tractatus* in its idiosyncratic style and composition lacks a clear argumentation structure and the goal is here to discuss in what way, if at all, extensionality should be taken as a thesis of the *Tractatus*. I argue that without connecting extensionality or its alleged thesis in sentence 5 to the operation $N(\bar{\xi})$ and the various forms of notations used in the *Tractatus*, there can be no clear understanding of the notion of extensionality that is shown in the *Tractatus*.

My thesis is that a proper understanding of extensionality in the *Tractatus* linked to $N(\bar{\xi})$ shows a demystification of logic. But that needs some qualification. The *Tractatus* does not fully expound what is now called first order logic. I do not aim to characterize the actual *Tractatus* logic in a formal way, but I retreat to the following formulation: the operation $N(\bar{\xi})$ demystifies the particular logic endorsed in the *Tractatus*. The demystification that I use here refers to the final passages of the *Tractatus* leading up to its famous final call for silence about the unspeakable. The mystical in those final passages is associated with the *feeling* of the world as a limited whole (T 6.45) and the unspeakable (T 6.522).

In this paper, I will approach extensionality in the *Tractatus* in three steps. Firstly, I will consider what $N(\bar{\xi})$ shows if it is taken as an operational device that generates a notation, rather than generating functions in their usual sense which is different from the one introduced in the *Tractatus*. Secondly, I will consider what role extensionality takes in the *Tractatus* according to Wittgenstein's own remarks.

In the third step I will reflect on extensionality in the *Tractatus* by comparing it to other uses of the term 'extensionality'. This will consist of one almost contemporary treatment explicitly mentioning Wittgenstein as the founder of the thesis of extensionality (Carnap's *Logical Syntax of Language*) and a recent retrospect on extensionality covering a long history of research into this topic (Quine's paper *Confessions of a Confirmed Extensionalist*).

[1] From here on this work while be referred to as the *Tractatus*. Reference to parts of the work will be given by the abbreviation T followed by a number according to the numbering system Wittgenstein introduces himself or just by a number when the context makes it clear that I am talking about the *Tractatus*.

The base for $N(\bar{\xi})$: elementary propositions

One central motivation for the *Tractatus* is analysis. The existence of a unique and complete analysis is very clearly stated in sentence 3.25. However, in this remark there is a curious vagueness of *what* there is one and only one analysis of. In German, Wittgenstein uses a singular possessive construction of 'the proposition' (*'des Satzes'*). Analytic Philosophy takes Russell's analysis of definite descriptions and Frege's analysis both of identity and the natural number as its paradigm cases of analysis. From this perspective one would rather expect to have a unique analysis of many different propositions. But Wittgenstein does not carry out any particular analysis as explicitly as Frege or Russell. Rather, and this gives the thesis of extensionality its special character in this context, the *Tractatus* radically scrutinizes not only the composition of the analysandum, but also the composition of the notation used for logical analysis itself. Characteristically, it is when he is talking about logical grammar or language and their signs that Wittgenstein's remarks in the *Tractatus* can be read as commenting on other thinkers' work (T 3.325, T 3.331, T 3.332, T 4.431, T 5.452). In those relatively rare cases the *Tractatus* can be seen as participating in an argumentative discussion rather than relying on the confessional but certainly not conventional style as the hallmark of the *Tractatus'* composition.

The notion of the elementary proposition is central to the understanding of the role of extensionality in the *Tractatus*. Elementary propositions are taken to be the simplest expressions asserting the existence of a state of affair (T 4.21). Here, 'exists' should be read in the sense of the 'what is the case' formulations in sentence 2 and, in particular, sentence 1.12, that states that the totality of facts also determines what is not the case. Elementary propositions are further characterized by their logical independence (T 5.134) and the impossibility of two elementary propositions contradicting each other (T 4.211). The existence of elementary propositions is a precondition for logical construction, that is, logical complexity, because they form the basis of the operation $N(\bar{\xi})$. This time, 'existence' of 'the existence of elementary proposition' requires a different reading than the one above.[2] Wittgenstein shuns any example of an elementary proposition and does not give any indication about a class of particular sentences to constitute elementary propositions. Rather, the application of logic determines which elementary propositions exist (T 5.557). I take this to be a very important aspect of elementary propositions because it means that the characterization of elementary propositions cannot and should not be taken as a positive test for being an

[2] The German text correspondence to this difference in meaning is the use of `Bestehen' and the corresponding 'Nicht-Bestehen' in sentence 2 and sentence 1.12 in contrast to the use of 'es gibt' in sentence 5.557.

elementary proposition and that it is fruitless to discuss candidates for elementary propositions. This can also be taken as pointing towards an interpretation that takes elementary propositions not so much as an observation about sentences or propositions at all, but rather as showing something about logic. What logic needs before any application are simple units that reflect the complete generality and unspecificness of states of affairs, of either being the case or not being the case.

The elementary propositions take a peculiar middle position between the opening part of the *Tractatus* with their reflection on the basic distinction between being the case or not being the case and the *feelings* associated with the mystical of the final passages. According to both the Ogden- and Pears/McGuiness-translations in sentence 4.411, Wittgenstein declared that the understanding of the general sentences depends *'palpably'* on the elementary propositions. However, the German word Wittgenstein uses and emphasizes in italics is '*fühlbar*'. It can be *felt* that the understanding of any proposition depends on the elementary propositions.

The sign of elementary propositions does not stand outside the realm of states of affairs themselves. Further evidence for this line of thought is sentence 4.221, where Wittgenstein declares that the ultimate goal of analysis is elementary propositions. Beyond that, logical analysis cannot go any further. Further decomposition does not yield further propositions but only names. So one way to look at the lack of examples and the lack of a positive test for elementary propositions is to take elementary propositions as the destination of logical analysis and accept sentences of ordinary and scientific talk as already being logically complex. Thus, logical analysis is not done when a preferred reading or interpretation of an ordinary sentence is presented in a logical notation, but when it is broken down to components that no longer show logical dependence on each other.

Again, analysis in this sense is not carried out in the *Tractatus*. What I take to be the main objective of sentence 5 and its comments according to the numbering system is to consider the methods for analysis. This is done by keeping open any decision about where to start with analysis and turn the direction around by asking what a notation that is completely void of any predetermination of application would look like and what successful analysis would ultimately lead to. That a sustainable method of analysis is actually something worth looking for is important for the Tractarian conception of philosophy, since critical inquiry into language is one task left for philosophy (T 4.0031), apart from delimiting the realm of science (T 4.113). However, if the method used for this critical inquiry is itself based on unfounded distinctions and stipulations it may be best to remain silent. That the *Tractatus* is not silent on the issues of logical notations is ample evidence that there is hope, in contrast to the realms of aesthetics and ethics. Although these may be felt to

be more important for the problems of our lives (T 6.52), they are beyond what can be talked about.

What Wittgenstein says in the *Tractatus* in sentence 5 about propositions is that a proposition is a truth-function of elementary propositions. The following interpretation of sentence 5 and extensionality in the *Tractatus* is guided by the Tractarian warnings that truth-functions are not material functions (T 5.44) and that they are not to be confused with operations (T 5.25). Heeding these warnings is what I call an *operational view* on extensionality in the *Tractatus*. What can be learned positively about the notion of truth-functions in the Tractarian sense is that they can be listed for a given number of elementary propositions (T 5.1, T5.101) and that they are *generated* by truth-operators (5.3). Later in the *Tractatus* the truth-operators are limited to only one, $N(\bar{\xi})$ (T 6.001). The next section focuses on how $N(\bar{\xi})$ can be thought of as a means to generate truth-functions.

Extensionality as operation on most general signs

It is in the sign of the elementary proposition and in the sign of the sentence in general constructed by the operation $N(\bar{\xi})$ that mirroring facts becomes possible without taking a point-of-view beyond or outside the world, the totality of states of affairs. The composition of the sign of the elementary proposition and the logical forms of the sentences in the Tractarian notation become clear because we chose to *write* it so and *give* the component signs *T* and *F*. As such, they are different from ordinary language sentences not composed for showing logical form but for other communicative means (T 4.002). *T* and *F* are mere convenience. Only two restrictions are important: the manifold of the composition of the elementary proposition must match the manifold of the division of being the case or not being the case, and the complex sentence must in principle be able to assert agreement or disagreement to all possibilities of obtaining and not obtaining that the elementary propositions require of which they are composed. The second restriction yields the listing of truth-functions for the case of two elementary sentences. Finally, in order to count as meaningful and expressible, the mark of the sentence must still offer different truth-possibilities, which in the notational variant means the appearance of both *T* and *F* in its sign.

To investigate what the operation $N(\bar{\xi})$ does for the *Tractatus'* view on logical notation, it is necessary to become acquainted with some of the instructions leading up to it. Some of these look like formulas or notation already, but by taking an *operational view*, I also mean to take these as abbreviations whose meanings are instructions on what to *do*. This procedure is textually supported in the *Tractatus*:

> 5.475 All that is required is that we should construct a system of signs with a particular number of dimensions—with a particular mathematical multiplicity.
>
> 5.476 It is clear that this is not a question of *a number of primitive ideas* that have to be signified, but rather of the expression of a rule.

The first case is the general term of a series of forms introduced in sentence 5.2522. Wittgenstein writes [*a, x, O'x*]. He tells the reader that this is a variable but not that it is a sign or symbol. The first term is to be understood as the beginning of a series, the second term is any member arbitrarily selected, and the last term is its immediate successor according to the application of an operation.

Additionally, Wittgenstein introduces schemata of combinations of the truth-possibilities of elementary propositions (T 4.31). Again, these are *not* declared signs. In order to get to the sign of a sentence these schemes need to be written down again with an additional row like in the following example:

p	q	
T	T	T
F	T	T
T	F	
F	F	T

This whole is declared a propositional sign (T 4.442). It is given the shorthand (*TTFT*)(*p,q*), but it is important to remember that this shorthand always depends on the scheme.

Finally, an operation is introduced in sentence 5.5: (– – – – – *T*)($\bar{\xi}$,...). This resembles the shorthand for the sentence sign above. It is given the shorthand $N(\bar{\xi})$.What that operation actually does, or rather, what one has to do to follow the rule expressed by $N(\bar{\xi})$ is the topic of the following paragraphs. In sentence 5.5 Wittgenstein declares that it negates all propositions in the right-hand pair of brackets.

The general form of the truth-function which is the general form of the proposition as well combines these instructions in the variable [$\bar{p}$, $\bar{\xi}$, $N(\bar{\xi})$].

Focusing on sentence 6.001 it becomes necessary to understand how the construction of logical notation comes about and what the content of the thesis of extensionality is within the *Tractatus*. Wittgenstein declares that each sentence is the result of successive application of the operation $N(\bar{\xi})$ on elementary propositions. He adds that with this, the transition from one sentence to another is also given (T 6.002).[3] But this cannot be taken as straightforwardas

[3] I ignore the puzzling introduction of the operation $\Omega'(\bar{\eta})$ as $[\bar{\xi}, N(\bar{\xi})]'(\bar{\eta})(= [\bar{\eta}, \bar{\xi}, N(\bar{\xi})])$. It apparently defies what is introduced in sentence 5.2522 about formal series, and corner brackets with just two comma

it might seem, because successive application of $N(\bar{\xi})$ to the result of a prior application does not yield the desired combinations of T and F as can be demonstrated for the case of two propositional variables for elementary propositions. The first suggestion from the *operational view* might be to exploit the resemblance of the operation $(- - - - - T)(\bar{\xi}, ...)$ with the shorthand for propositional signs and assume that the application of $N(\bar{\xi})$ consists of writing down the appropriate scheme for the number of propositions and filling in an F in the lowermost row. In shorthand: (TTTF)(p,q). But just that would rule out *successive* application.

Additionally taking negation as switching the *T*s and *F*s used in *writing down* the schemes of elementary sentences we get the following progression:

(1) $N((TF)(p),(TF)(q)) = (FFFT)(p,q)$ $\quad plq$

(2) $N((FFFT)(p,q)) = (TTTF)(p,q)$ $\quad p \vee q$

(3) $N((TTTF)(p,q)) = (FFFT)(p,q)$ $\quad plq$

Any further application of $N(\bar{\xi})$ would obviously only reiterate this situation. Taking into account the application of $N(\bar{\xi})$ to only a single elementary proposition we get this:

(4) $N((TF)(p)) = (FT)(p)$ $\quad \sim p$

(5) $N((FT)(p)) = (TF)(p)$ $\quad p$

Interestingly enough, this procedure yields two sets of functionally complete connectives from a modern perspective: {*I*} and {∨,~}.[4] However, the point I intend to make is that the operation can be taken as instructions on how to construct all the listable truth-combinations of sentence 5.101 as truth-functions in the Tractarian sense, not by combination of connectives associated with one or some of the truth-functions in the modern sense of being a function from a set of truth-values to truth-values. If, in the case of two propositional values, the introduction of *plq* by the associated function $\{F, F\} \longrightarrow T$ and *F* otherwise, is taken as the whole point of $N(\bar{\xi})$, then the conception collapses back into a functional instead of a operational view about logical form in the *Tractatus*.

Another possibility is to take the schemes in sentence 4.31 themselves, take them as $\bar{p}$ and successively operate on by taking neighboring rows as consecutive entries as $\bar{\xi}$ and recording 'T' only in case two 'F's are each other's neighbor:

separated entries are not used anywhere else in the *Tractatus*.

[4] There is a potential source of confusion here over the sign that is associated with the function ($\{F, F\} \longrightarrow T$ and *F* otherwise), by the functional view. In the *Tractatus,* Wittgenstein introduces the *I* as 'neither p, nor q' (T5.1311). But the sign *I* is now called Sheffer stroke and more commonly associated with the function ($\{W, W\} \longrightarrow F$ and T otherwise). I will stick to the *Tractatus* convention when discussing *I* and take *plq* as a whole as a name for the truth function in the Tractarian sense (FFFT)(p,q).

P	Q	$(p,q)_1$	$(p,q)_2$	$(p,q)_3$	$(p,q)_{4=1}$
T	T	F	F	T	F
F	T	F	F	T	F
T	F	F	T	F	F
F	F	T	F	F	T
$\bar{p}$		$\bar{\xi}_2$			
	$\bar{\xi}_1$		$\bar{\xi}_3$		

Again, the procedure terminates too early and does not achieve what $N(\bar{\xi})$ is supposed to.

Another dead end is to start with the schemes and then allow recombination of any produced column with the signs of the elementary sentences in the proper manifold that is then determined by the fact that the other function $(----)(p,q)$ already requires four entries of 'T' or 'F'. It is at least not enough in its own right, but it can be augmented to generate all truth-functions. The following short-hands of the generation start with the restriction of recombination with the elementary propositions which is afterwards lifted.

(6) $N((TFTF)(p), (TTFF)(q)) = (FFFT)(p,q)$ $(p,q)_1$

This first step already calls for a pause, though. Wittgenstein repeatedly associates this first step of applying $N(\bar{\xi})$ with the *Principia* notation '$\sim p.\sim q$' (T 5.101, 5.1311). But in the proper notation it is revealed that neither $\sim p$ nor $\sim q$ are components in the proper understanding of complexity of the sentence sign: $(FFFT)(p,q)$. It is important that this is the *first* step in logical composition on the basis of two elementary propositions. There can be no other. It is possible that this result is reached again by more complex composition, but no other truth-function can be reached before that. What the composition of '$\sim p.\sim q$' suggests is that there is a combination of two smaller parts which are in turn not simple. This is revealed to be misleading by Tractarian lights.

What is shown is that from the Tractarian perspective, the *Principia* notation actually suffers from the same deficits that natural language is ridden with. The way *this* notation is concatenated and treated makes it look as if there were more structure than there actually is. While it can be used to show that what seems to be the logical form of a proposition is not its actual one (Russell's merit, according to 4.0031), the notation itself does not live up to the standard of revealing instead of expressing the real logical form. That argument leaves a lacuna as to whether reformulation of the *Principia* system into one with I as the sole primitive would be acceptable from the viewpoint of the *Tractatus*. The remarks in 5.1311 do point in this direction. However, this cannot be addressed without evoking the requirements of autarky of logical matters and avoidance of hierarchy of logical propositions. It is not shown by the operative perspective on truth-combinations alone.

Successive application from (6) then yields:

(7) $N((FFFT)(p,q),(TFTF)(p)) = (FTFF)(p,q)$ $(p,q)_2$

(8) $N((FFFT)(p,q),(TTTF)(q)) = (FFTF)(p,q)$ $(p,q)_3$

(9) $N((FTFF)(p,q),(TFTF)(p)) = (FFFT)(p,q)$ (6)

(10) $N((FTFF)(p,q),(TTFF)(q)) = (FFTT)(p,q)$ $(p,q)_4$

(11) $N((FFTF)(p,q),(TFTF)(p)) = (FTFT)(p,q)$ $(p,q)_5$

(12) $N((FFTF)(p,q),(TTFF)(q)) = (FFFT)(p,q)$ (6)

(13) $N((FFTT)(p,q),(TFTF)(p)) = (FTFF)(p,q)$ (7)

(14) $N((FFTT)(p,q),(TTFF)(q)) = (FFFF)(p,q)$ $(p,q)_6$

(15) $N((FTFT)(p,q),(TFTF)(p)) = (FFFF)(p,q)$ (14)

(16) $N((FTFT)(p,q),(TTFF)(q)) = (FFTF)(p,q)$ (8)

(17) $N((FFFF)(p,q),(TFTF)(p)) = (FFFT)(p,q)$ (6)

(18) $N((FFFF)(p,q),(TTFF)(q)) = (FFTT)(p,q)$ (10)

At this point, the operational procedures terminate. However, others can be generated by taking the ones produced and using them as single values for $N(\bar{\xi})$. This must exclude the initial columns of p and q though, else this would confuse $N(p)$, which is $(FT)(p)$ and not $(FTFT)(p)$.

(19) $N((p,q)_1) = (TTTT)(p,q)$ $(p,q)_7$

(20) $N((p,q)_2) = (TFTT)(p,q)$ $(p,q)_8$

(21) $N((p,q)_3) = (TTFT)(p,q)$ $(p,q)_9$

(22) $N((p,q)_4) = (TTFF)(p,q)$ $(p,q)_{10}$

(23) $N((p,q)_5) = (TFTF)(p,q)$ $(p,q)_{11}$

(24) $N((p,q)_6) = (TTTT)(p,q)$ $(p,q)_{12}$

Only four binary truth functions are missing. These can now be generated in numerous ways. One example for each are the following shorthands, again with the reservation that these should really be thought about as introduced above:

(25) $N((p,q)_3, (p,q)_5) = (TFFF)(p,q)$ $(p,q)_{13}$

(26) $N((p,q)_{13}, = (FTTT)(p,q)$ $(p,q)_{14}$

(27) $N((p,q)_2, (p,q)_3) = (TFFT)(p,q)$ $(p,q)_{15}$

(28) $N((p,q)_{15}) = (FTTF)(p,q)$ $(p,q)_{16}$

The method must now conclude, because all 16 possible combinations of writing T and F have been passed through. This is not enough to demonstrate that this method works for all bases of elementary propositions and a demonstration of that cannot rely on functional completeness of the truth-function associated with the Sheffer stroke ($\{F , F\} \dashrightarrow T$ and F otherwise). I do

not have such a demonstration, but there are some things that can be pointed out for the first application of $N(\bar{\xi})$ for a base of three elementary propositions:

p	Q	R	*N(p,q,r)*
T	T	T	F
F	T	T	F
T	F	T	F
T	T	F	F
F	F	T	F
F	T	F	F
T	F	F	F
F	F	F	T

This makes it clear that, in principle, the method is still available to check each line and write *T* in each line there are only *F* s in the way it was introduced above. So basically, the notational practice of the *Tractatus* comes down to checking a line occurrence of the same entry. Which, significantly, is the same way we determine in this procedure both tautologies and contradictions but, in that case, by columns.

In the tautologies and in the contradictions the compositions of the symbol collapse because the manifold that is required for presenting a state of affairs is lost. This does not mean that there is no complexity in the sign (T 4.4661). It is important that it is not lost, for this keeps them open for application of the operation $N(\bar{\xi})$. Even though the logical sentences do not enable representation, they are still within the reach of the construction of the meaningful sentences by $N(\bar{\xi})$.

The Tractarian distinction of saying and showing has received enormous attention but some of the many entangled problems in this context of this distinction cannot be untangled unless writing or writing down as an action does not also receive some attention in the *Tractatus*. The logical form of the sentence is something that cannot be said as it must be shown. But the structure of the notation is not in this way beyond the reach of analysis. It shows itself immediately because we *choose* to set it up this way. In that, we are free to declare what the components of construction are and it is clear from the manipulation of signs, in the *Tractatus* the single truth operation $N(\bar{\xi})$, which are the units of this manipulations because they are pointed out by declaring them so. In this way, we fix the relevant parts of the state of affair that is the notation of a sentence and in particular, an elementary proposition.

How $N(\bar{\xi})$ serves the *Tractatus* view on logical notation

In this part I will focus on what is claimed in the *Tractatus* about extensionality introduced by the generation of a notation designed to show its

logical form. From the more conventional perspective of modern logical expression and form, tutored by assuming distinction of syntax and semantics, it looks as though Wittgenstein supplied merely a method of truth tables as a semantics and the reduction of all truth table combinations to one operation. However, from the Tractarian perspective much more is at play, and I will present these under the general label of demystification of logic.

The first achievement from the Tractarian perspective is to eradicate the distinction between axioms and theorems in a logical system. This has several advantages from the Tractarian view on logic in general, and the laudation of this merit finds expression in several formulations.

One aspect of this is that composition by $N(\bar{\xi})$ obliterates any hierarchy among logical sentences. Wittgenstein does not want to convince the readers of the *Tractatus* that there is a special class of logical sentences, the axioms, from which other logical sentences are derived. This would immediately call into question where these axioms come from, and why they are these and not others. Furthermore, this very procedure of setting up a logic would suggest that there are many sentences of logic that say different things, but all logical sentences say nothing, according to sentence 5.43.

This is closely connected to the demand for logic expressed in the *Tractatus* claiming logic must take care of itself (T 5.473). A procedure that takes a class of axioms to start with to produce more logical sentences needs some justification as to why these were logical to begin with. But that would have to be an external justification. The operation $N(\bar{\xi})$, in contrast, as the manifestation of logical composition, does take a start somewhere, either a selection of elementary sentences or possibly their indefinite whole, but these are just the simplest possible units of independent presentation of states of affairs. The fact that one can determine which logical constructions are logical sentences with this meager basis relying only on the reflection of the possibility of being the case and not being the case is important. In the context of the *Tractatus* the proper notation must make it possible to *see* them in the sign of the sentence (T 6.122).

With the operation $N(\bar{\xi})$ it is meant to show that construction of a logical notation that 'takes care' of itself is not only free of outside justification but also reduction to the most simple procedure of discovering the logical sentences is possible. It is not the rules of inference that are most intuitive and simple suggesting more and more complex sentences of logic, but the procedures of checking rows and columns for the proper manifold. It is in this way that rules of inference in a proper notation would become superfluous (T 5.132).

Both requests, autarky and freedom of hierarchy reflect on a written notation as provisions to make analysis possible. Ordinary language is already in proper logical order (T 5.5563), but ordinary language is not meant to show the formal properties of their construction but is made for other communicative

means. Only a constructed language can show the formal properties (T 5.556). This construction, which is a doing, must be the most simple thing (T 5.4541). The operational approach presented as a construction of notation is just as simple as the very demand to distinguish between the possibility to obtain or not to obtain.

Another aspect of the demystification of logic in the *Tractatus* is that tautologies and contradictions receive equal treatment in the notation. In principle, this reflects that logic is not concerned with the truth or falsity of sentences. Since both the forms of tautologies and contradictions are set out by the same mechanism and show the same defect from the perspective of presentation, namely the lack of sufficient manifold in their notation.

In sentence 6.1202, Wittgenstein observes that contradictions might as well be put to the same use as the tautologies. Since logic needs to be put to all uses, contradictions serve as limits to meaningful notation just as much as tautologies do.

Finally, the operation $N(\bar{\xi})$ and extensionality account for the logical connectives, particularly negation, the status of which is questioned for example in sentence 5.512:

5.512 '*p*' is true if '$\sim p$' is false. Therefore, in the proposition '$\sim p$' when it is true, '*p*' is a false proposition. How can the stroke '$\sim$'make it agree with reality?

The answer in the *Tractatus* is that the stroke does not do anything. The only thing that can be the negation, which makes agreement with reality an option, is that which is common to all the variants in *this* notation that are an equivalent with '*p*': '$\sim\sim\sim p$', $\sim p \vee \sim p$' which in turn is that no line of their associated truth function has anything in common with the sign of (*WF*)(*p*) which is the mark of the negative (of which there is only one) of a sentence to another (T 5.513). This is extended to all the connectives and declared a fundamental idea:

4.0312: My fundamental idea is that the 'logical constants' are not representatives; that there can be no representatives of the logic of facts.

In effect, in a proper notation, the signs of a logical notation just *are* most general facts that mirror states of affairs. And they are so, because they are made.

Tolerance and rigor: Carnap and Quine on extensionality

The final step to approach extensionality in the *Tractatus* is to compare it with other cases where the term is used.

Carnap's *Logical Syntax of Language* mentions Wittgenstein and Russell as adhering to a thesis of extensionality (CARNAP, Logical Syntax of Language,

p. 245). He also cites his own *Logischer Aufbau der Welt.* What all these sources neglect, according to Carnap, is that there is not only one language but several. He aims this criticism particularly at Wittgenstein who is accused of using '*the* language' with a definite article throughout the *Tractatus*.

We have seen that Wittgenstein maintains that there is only one final analysis and that there is the goal of finding the one notation that makes everything clear and easy. However, Wittgenstein acknowledges the possibility of different symbolic systems in sentence 4.5:

> 4.5 It now seems possible to give the most general propositional form: that is, to give a description of the propositions of any sign-language whatsoever in such a way that every possible sense can be expressed by a symbol satisfying the description, and every symbol satisfying the description can express a sense, provided that the meanings of the names are suitably chosen.

Wittgenstein adds that this description of the proposition of any such language can only be the most general form. What Wittgenstein aims at is on a different level than Carnap's thesis of extensionality which is a claim about the translatability of different languages into each other. Wittgenstein is trying to show something about the possibility of constructing any notation or sign language, and that requires not only making the distinction of being the case or not being the case, but also at most that distinction. The translation of languages must rely on the claim that at least something is the case: some parts of these languages correspond to each other. But this makes it clear that this form of extensionality already presupposes the one Wittgenstein has in mind.

Carnap takes the possibility of non-extensional languages seriously and retreats to a statement he considers more cautious than the thesis of extensionality associated with sentence 5 in the *Tractatus*, the claim that all propositions are truth-functions of elementary propositions.

The aim of *Logical Syntax* is very similar to what has been claimed about the motivation of the *Tractatus*. Carnap claims that the scientific work of the philosopher is logical analysis (CARNAP, 1937, p. 13). However, Carnap's approach is to give a method for talking about the sentences of logic and to express the exact manner the findings of logical analysis. While Wittgenstein is interested in letting logical notation show its formal properties from within, Carnap takes the route of formulating a metalogic.

Finally, Tractarian extensionality is similar yet subtly different from the doctrine of extensionalism Quine espoused throughout his career. I shall take Quine's self-commenting and self-summary of *Confessions of a Confirmed Extensionalist* as a convenient way to look at some of the many aspects that are linked to extensionality in Quine's thinking. Famously, and as Quine points out himself in the opening paragraph, extensionalism has been a phi-

losophical doctrine or *policy* that Quine held on through his whole career. Extensionalism is declared "a predilection for extensional theories." (QUINE, 2004, p. 329). But more important than this is the definition of extensionality itself: [A]n expression is extensional if replacement of its component expressions are by coextensive expressions always yields a coextensive whole (QUINE, 2009).

There is a similar focus on notation, particularly in the notion of *semantic ascent*, the "strategy to talk about expressions"(QUINE, 2004, p. 337) instead of talk about things with unclear identity conditions, like properties or meaning. In this way, there appears to be similarity between Quine and Wittgenstein's approach to logical notation. However, in getting at the notion of coextensive expressions we see how different the approaches in fact are. Quine discusses the coextension of three sorts of expressions: closed sentences; predicates, general terms, and open sentences; as well as singular terms. Much to the point, what predicates, open sentences, and general terms are only comes together once there is already a logical theory: "They are what the open letters in quantification theory stand for. Open sentences are the most graphic of the three renderings." (QUINE, 2004, Ibid., p. 329). Unlike Wittgenstein's Tractarian conviction that proper logical notation must be guided by the consideration of extensionality that is itself manifested in the signs of the logical notation, for Quine logical notation comes first.

As to the usefulness of extensionality Quine cites the 'clarity and convenience' that come with the possibility of interchanging coextensive components *salva veritate*. This in turn gets so much emphasize that Quine adds: "I doubt that I have ever fully understood anything that I could not explain in extensional language" (QUINE, 2004, p. 331).

This does, in a way, put Quine's view on extensionality closer to Wittgenstein's than Carnap's. Carnap's version of the thesis of extensionality that accepts intensional languages by their own right but claims they are open for translation into an extensional language presupposes that those intensional languages can be understood. Both Quine and Wittgenstein take the stand that outside the extensional, there is nothing to understand.

The motivations for both Quine and Wittgenstein were apparently also very close. We see Wittgenstein complaining about the use of 'words' in the introduction of definitions and basic laws of *Principia* in sentence 5.452 and, likewise, Quine describes what bounded his admiration for the *Principia*:

> My admiration was not quite unbounded. It was bound by the explanations in prose that were preposed and interposed as explanatory chapters and in briefer bits among the expanses of symbols (QUINE, 2004, p. 332).

It is also clear that there is a similar concern about the primitives of the foundations used in the notational language. Quine describes his two stages

of improving the *Principia* by using first individuals, classes and sequences in his dissertation, and finally just class inclusion and class abstraction in 1937, instead of propositional functions. Quine holds propositional functions of the *Principia* to be identifiable with propositions in the case of application to one variable, and to relations in the case of more variables. However, their lack of a principle of individuation makes them unclear.

Finally, then, one can say that even though the motivations for 'extensional' reform of logical notation in both Wittgenstein and Quine are similar, their ultimate concern is different. Wittgenstein wants an operation that with its clarity and simplicity makes the puzzling questions about notation vanish, while Quine constructs a formal foundation that adheres to the standard of 'no entity without identity'. For Quine, extensionality expresses the standard of clarity and simplicity in interchangeability and identity, questions about what there is. For the early Wittgenstein, it is the simplicity of applying a most simple rule to follow, a question about what to do.

References

BLACK, M. *A Companion to Wittgenstein's 'Tractatus'*, first published. Cambridge: University Press, 1964, reprinted Ithaca: Cornell University Press, 1992.

CARNAP, R. *Logical Syntax of Language*, first published by P. Kegan, translated by A. Smeaton, London: Trench, Trubner & Co., 1937, reprinted London: Routledge, 2001.

FRASCOLLA P. *Understanding Wittgenstein's Tractatus*. London: Routledge, 2007.

QUINE, W.V. "Confessions of a Confirmed Extensionalist." In: FLOYD, J.; S. Shieh. *Future Pasts*. New York: Oxford University Press, 2001, reprinted in *Quintessence:* basic readings from the philosophy of W.V. Edited by R. F. Gibson, Jr., 329 – 337. Cambridge (MA), London: The Belknap Press of Harvard University Press, 2004.

ROSENBERG, J. F. "Intentionality and Self in the Tractatus." *Noûs*, v. 2, n. 4 1968, p. 341 – 358.

WITTGENSTEIN, L. *Tractatus Logico-Philosophicus*, translated from the German by C.K. Ogden, with an Introduction by Bertrand Russell. London: Boston and Henley: Routledge & Kegan Paul 1981. Online: K.C. Clement: Side-by-side-by-side edition, version 0.42 (January 05, 2015), containing the original German, alongside both the Ogden/Ramsey, and Pears/McGuiness English translations., http://people.umass.edu/phil335-klement-2/tlp/tlp.html.

ARGUMENTOS
Revista de Filosofia

Diogo Henrique Bispo Dias*

Carnap's Principle of Tolerance and logical pluralism

RESUMO

Pluralismo lógico é a tese de que há mais de uma lógica adequada. Diversos autores apontam Carnap como um dos precursores do pluralismo lógico. Mais que isso, afirmam que o Princípio de Tolerância consiste em uma das primeiras formulações explícitas de um pluralismo lógico. Não obstante, há poucas e esparsas investigações detalhadas para avaliar se o Princípio de Tolerância implica necessariamente em um pluralismo lógico e, caso implique, de qual tipo. O objetivo deste artigo é analisar o Princípio de Tolerância, bem como o contexto no qual tal princípio está inserido e, por fim, investigar qual a relação entre esse princípio e o pluralismo lógico.

Palavras-chave: Carnap; princípio de tolerância; logical syntax; pluralismo lógico.

ABSTRACT

Logical pluralism is the claim that there is more than one adequate logic. Many authors consider Carnap as one of the forerunners of logical pluralism. More than that, they claim that Carnap's Principle of Tolerance consists in one of the first explicit formulations a logical pluralism. Nonetheless, there is little detailed investigation to evaluate if the Principle of Tolerance necessarily implies a logical pluralism, and if so, of which kind. The aim of this paper is to analyze the Principle of Tolerance, as well as its context, and to investigate the relation between such principle and logical pluralism.

Keywords: Carnap; Principle of Tolerance; logical syntax; logical pluralism.

* University of São Paulo – USP, Doctoral Student. Email: diogo.bispo.dias@gmail.com

Introduction

This paper has a double aim. On the one hand, it intends to analyze Carnap's Principle of Tolerance presented on his book *Logical Syntax of Language*. For this, we will investigate not only this work, but also the influence of other thinkers for Carnap's thought. On the other hand, this paper intends to answer the following question, namely: what is the relation between Carnap's Principle of Tolerance and a possible logical pluralism? This is an important task since, although many scholars consider Carnap as one of the forerunners of logical pluralism, there is little detailed investigation to evaluate if the Principle of Tolerance necessarily implies a logical pluralism, and if so, of which kind.

Logical syntax of language and the Principle of Tolerance

Let us begin with the Principle of Tolerance. This is formulated for the first time in *Logical Syntax of Language*. In general terms, the main goal of the book is

> to provide a system of concepts, a language, by the help of which the results of logical analysis will be exactly formulable. Philosophy is to be replaced by the logic of science – that is to say, by the logical analysis of the concepts and sentences of the sciences, for the logic of science is nothing other than the logical syntax of the language of science. (CARNAP, 1937, p. viii).

With the formulation of a general syntax, applicable to any language, Carnap intends to present a solution to many philosophical problems. In fact, the idea of a general syntax was meant to replace philosophy itself[1].

Nonetheless, there is a particular problem that occupies a central position in this book, namely: the discussion between formalism, intuitionism and logicism regarding the foundation of mathematics.

It is precisely in this context that the Principle of Tolerance emerges[2]. Once a syntactical metalanguage is formulated, it is possible to see that the three proposed solutions consist merely in formulations of different languages. In other words, from this perspective, logicism, formalism and intuitionism consist of three different ways of formulating a language, i.e., of stipulating a set of symbols together with some rules for their manipulation. In this level

[1] We will limit ourselves to evidence the aspects of *Logical Syntax* that allows the formulation of the Principle of Tolerance. We will not investigate any specific problems of *Logical Syntax, n*or its application to other philosophical problems.

[2] Carnap claims, years later, that "it might perhaps be called more exactly the 'principle of the conventionality of language forms'". (CARNAP, 1963, p. 5).

there is no need for an external justification for such formulation. Thus, the Principle of Tolerance consists in affirming that

> *In logic, there are no morals*. Everyone is at liberty to build up his own logic, i.e. his own form of language, as he wishes. All that is required of him is that, if he wishes to discuss it, he must state his methods clearly, and give syntactical rules instead of philosophical arguments. (CARNAP, 1937, p. 52).

Therefore, this principle does not offer a solution to the problem regarding the foundation of mathematics, but represents a dissolution of the problem. One must understand the proposed solutions as suggestions to build a language. After that, what remains is to investigate and evaluate the consequences that follow from each language.

In a schematic way, Carnap's aim with his general syntax and his Principle of Tolerance is to put an end to any discussion regarding the justification, or 'truth' of a logic. According to the German philosopher, it is precisely this pursuit for justification that, on the one hand, precludes investigations on countless languages different than classical logic and, on the other, creates several pseudo-problems on the subject. In this – and only in this – sense it is possible to understand Carnap's project as a solution for the foundations of mathematics.

The solution, therefore, does not consist of a synthesis between formalism, logicism and intuitionism, but of the possibility to establish a common framework on which each proposal can be formulated and its advantages – as well as its disadvantages – can be explored.

From this change of perspective, several questions about the foundations of mathematics must be reformulated. Take, for instance, the debate on the possibility of using impredicative definitions. Carnap asserts that

> The proper way of framing the question is not 'Are […] impredicative symbols admissible?', for, since there are no morals in logic […] what meaning can 'admissible' have here? The problem can only be expressed in this way: 'How shall we construct a particular language? Shall we admit symbols of this kind or not? And what are the consequences of either procedure?' It is therefore a question of choosing a form of language – that is, of the establishment of rules of syntax and of the investigation of the consequences of these. (CARNAP, 1937, p. 164).

It's important to highlight that the notion of logic in the *Logical Syntax* is much more comprehensive than today's concept of logic. For Carnap, tolerance towards logic means tolerance with respect to the language adopted. The languages may contain different inferential apparatus, such as arithmetic and type theory. All different kinds of formulations are allowed, since the symbols introduced and the syntactical rules for their manipulation are explicitly presented. Put in other words, a language or a linguistic framework

is syntactically specified by its formation and transformation rules. So far, this is the common procedure nowadays. Nonetheless, for Carnap, a language can possess formal and empirical components. Thus, both logical and physical rules are needed. In Carnap's terminology, we have the L-rules – the logical rules –, and the P-rules – the physical rules –; the L-consequences, and the P-consequences.

Although there is a precise distinction between L-sentences and P-sentences once a language is formulated, both kinds of sentences are open to revision. Nevertheless, when a P-rule is reformulated, this is done within the same language. Thus, this is only a reformulation of the empirical sentences formulated in the language. Reformulating an L-rule, on the other hand, represents changing the language, since we are changing the behavior of its symbols. Moreover, there is no absolute extra linguistic division between these rules. Such division is only possible after the establishment of a language. Thus it is possible in principle that a rule is logical in a language, and empirical in another one[3].

The very notion of syntax in this context is different from its present meaning. Carnap "developed the idea of the logical syntax of a language as the pure analytical theory of the structure of its expression." (CARNAP, 1963, p. 53). Some concepts present on his general syntax, such as the concept of analyticity, would be considered as semantic concepts today. In this sense, it is curious to note that Carnap has named this project, on moments prior to its publication, as metalogic and even as semantics.

After presenting such rules,

> The investigation will not be limited to the mathematico-logical mart of the language [...] but will be essentially concerned also with synthetic, empirical sentences. The latter, so-called "real" sentences, constitute the core of science; the mathematico-logical sentences are analytic, with no real content, and are merely formal auxiliaries. (CARNAP, 1937, p. xiv).

In this quotation one can see an essential point. The distinction between the empirical and conventional sentences is established by the distinction between analytical and synthetic sentences. Once again, this distinction is no longer absolute. To formalize a language is, at the same time, to distinguish analytical from synthetic sentences. And this distinction is fundamental, for "all of logic including mathematics, considered from the point of view of the total language, is [...] no more than an auxiliary calculus for dealing with synthetic statements". (CARNAP, 1935/1953, p. 127).

[3] For an example, cf. FRIEDMAN, 1999, p. 85.

Therefore it is the concept of analyticity that assures that both logic and mathematics do not say anything about the word, although they may be used as auxiliaries to analyze empirical sentences[4].

As seen, according to the Principle of Tolerance one can develop different kinds of languages. The next question is: How to formulate different linguistic frameworks and why? The answer is found once it is acknowledge the auxiliary role of logic and mathematics inside a linguistic framework. According to Carnap:

> if we regard interpreted mathematics as an instrument of deduction within the field of empirical knowledge rather than as a system of information, then many of the controversial problems are recognized as being questions not of truth but of technical expedience. The question is: Which form of the mathematical system is technically more suitable for the purpose mentioned? Which one provides the greatest safety? If we compare, e.g., the systems of classical mathematics and of intuistionistic mathematics, we find that the first is much more simples and technically more efficient, while the second is more safe from surprising occurrences, e.g., contradictions. (CARNAP, 1939, p. 50).

So, this choice is determined by a purely pragmatic question, namely: which is the purpose of this formulation? In this way, questions regarding the validity of a given argument can be understood in two senses. First, we can look at it as an internal question. This means that we choose a logic, formalize the argument, and then we evaluate its validity. Another option is to understand this question as an external one. We may ask: Should this argument be interpreted in Language L or L*? Note that, in both cases, there's no absolute answer regarding the validity of a given argument.

Influences on Carnap

Carnap states in his autobiography that the ideas presented in *Logical Syntax* occurred to him during a sleepless night:

> the whole theory of language structure and its possible applications in philosophy came to me like a vision during a sleepless night in January 1931, when I was ill. On the following day, still in bed with a fever, I wrote down my ideas on forty-four pages under the title "Attempt at a Metalogic". These shorthand notes were the first version of my book *Logical Syntax of Language*. (AWODEY, CARUS, 2007, p. 24).

This does not mean, however, that Carnap developed these ideas in isolation. It is well known that Carnap was influenced not only by the other

[4] There is a long dispute regarding whether Carnap's project and specially his concept of analyticity is subject to Gödel's incompleteness theorems. Since this problem is tangential to the subject of this paper, we will not discuss about it. For an inquiry on this debate, (FRIEDMAN, 1999 and RICKETTS, 1994).

members of the Vienna Circle, but also by other authors working on logic and foundations of mathematics. Thus, this section aims to analyze these authors' influence on Carnap's thought.

As said before, Carnap proposes not a synthesis, but an overcoming of the debate between formalism, logicism and intuitionism. Nonetheless, Carnap's proposal is very different from these three doctrines. The introduction of the numbers and the induction principle as primitive symbols and axiom seems to indicate that logicism was rejected. Similarly, the use of non-constructive methods suggests that intuitionism was abandoned. Furthermore, the presence of infinitary reasoning indicates that formalism was also ruled out. This allows us to question which aspects are preserved from these three schools in *Logical Syntax.*

Starting with formalism, it is clear that, despite the divergences noted above, Carnap retain Frege's idea that there is no need for a foundation of logic, for every rational discourse presupposes logic. In addition to that he also incorporates Fregean anti-empirism regarding mathematics. And this leads him to accept the logicist's thesis that mathematics can be reduced to logic. Thus, mathematics is also analytic. The main differences are in the rejection of a universal conception of logic, as well as in the acceptance of the axiomatic method and the distinction between language and metalanguage.

It is precisely this acceptance above that Carnap retains from formalism. Languages I and II formulated on *Logical Syntax* are presented in an axiomatic way. Besides that, Tarski's and Hilbert's investigations on metalanguage allow Carnap to formulate a general syntax, i.e., the possibility of discussing on the metalanguage the general features of any language. Tarski himself, while visiting the Vienna Circle in January of 1930, was responsible to personally stress to Carnap that

> concepts used in logical investigations, e.g., the consistency of axioms, the probability of theorems in a deductive system, and the like, are to be expressed not in the language of axioms (later to be called the object--language), but in the meta-mathematical language (later called the metalanguage). (CARNAP, 1963, p. 30).

On the other hand, after Gödel's incompleteness theorems, Carnap abandons any formalist pretention to present a finitary consistent proof for classical mathematics. Thus, he rejects the thesis that consistency proofs of axiomatic theories assure the existence of this theory's objects. Formalism also holds the universality of logic, something which Carnap rejects, as seen. Furthermore, the languages investigated in *Logical Syntax* differ from the formalist project in the sense that they are not limited in formalizing mathematics, but are also used as formalization of science and allow the use of non-decidable concepts.

With respect to intuitionism, Carnap acknowledges that the lecture given by Brower to the Vienna Circle in 1928 had a strong influence on him. Carnap himself recognizes that influenced by Brower he had "a strong inclination toward a constructivist conception." (CARNAP, 1963, p. 49). Thus, language I is formulated based only on primitive recursive arithmetic which in principle only uses constructive reasoning and concepts. Furthermore, Carnap asserts several times that constructive methods are safer than others.

It is curious to note however that it seems that from the three doctrines, intuitionism is the least influent for *Logical Syntax*'s project. Although language I is constructive, the fact that it is formulated axiomatically and that the intuitionist concept of continuum cannot be formalize on it suggests that Brower himself would reject it. The very use of a classical metalanguage to formalize language I would also be rejected by intuitionism. In addition to that, one of the main criticisms from intuitionism towards classical logic – that classical treatment of logical connectives is incoherent – becomes unfounded in Carnap's project.

Other authors were also fundamental to the development of Carnap's thinking. The first version of *Logical Syntax* – called *Attempt at a Metalogic* – did not contain the Principle of Tolerance. This appears, partly, as an answer to the criticism made by Gödel to the first drafts of this work. In this first version, Carnap tried to define a general concept of analyticity. In personal communication, Gödel showed him that such definition was flawed[5]. Carnap acknowledges, after the works of Tarski and Gödel, that such concept was to be defined in a metalanguage. But at that time, he thought that by using "Godel's method of arithmetizing the metalanguage in the object language, [...] one could now get by with only a single language after all." (AWODEY, CARUS, 2009, p. 93). Gödel was responsible for showing Carnap that this method was subject to his incompleteness results. So it was necessary not only a metalogic with a greater power of expression than the object language, but also a hierarchy of such metalanguages. Thus, given a language L, the concept of analytic-in-L must be formulated in the metalanguage of L. Therefore, analyticity is always defined relatively to a language. From these considerations, Carnap extends this notion of linguistic relativity and starts to claim that many linguistic frameworks are legitimate.

Once the legitimacy of several languages is recognized, the choice among them becomes a merely pragmatic issue. Here we can clearly see the influence of Poincaré: "What, then, are we to think of the question: Is Euclidian geometry true? It has no meaning [...]. One geometry cannot be more true than another; it can only be more convenient." (POINCARÉ, 1905, p. 50). Thus, Carnap extends Poincaré conventionalism to logic itself.

[5] For a detailed exposition of this discussion cf. AWODEY, CARUS, 2007 and AWODEY, CARUS, 2009.

Lastly, we have to acknowledge Wittgenstein's importance. *Tractatus* exerted a strong influence on the Vienna Circle. In particular it is fundamental to highlight that Carnap accepts Wittgenstein's position that logical truth are tautologies. Thus, logic does not talk at all about the world. Now, since mathematics can be reduced to logic – something that Carnap incorporated from logicism –, it results that mathematical truths are also tautological. But in order to achieve *Logical Syntax'*s project, it was necessary to depart from some fundamental ideas on *Tractatus:*

> the members of the Circle, in contrast with Wittgenstein, came to the conclusion that it is possible to speak about language and, in particular, about the structure of linguistic expressions. On the basis of this conception, I developed the idea of the logical syntax of a language as the purely analytic theory of the structure of its expression. My way of thinking was influenced chiefly by the investigations of Hilbert and Tarski in metamathematics. (CARNAP, 1963, p. 53).

In addition to that, the *Tractatus* also holds, in certain sense, the absolute character of logic. Therefore, not a general theory of language, nor the principle of tolerance could be formulated in the *Tractatus* terms. Here it is important to clarify some points regarding the posterior development of Wittgenstein's thought. A few years after the publication of *Tractatus*, and before the publication of *Logical Syntax*, Wittgenstein wrote about the possibility of formalizing freely different languages. Nonetheless, this change did not exert any influence on the formulation of the Principle of Tolerance. Firstly, Carnap did not read those texts. When Schlick was reading one of the drafts of *Logical Syntax*, he wrote a letter to Carnap warning him that Wittgenstein was also developing something along the same line. Thus, in the first version of *Logical Syntax*'s Foreword, Carnap affirms that:

> A propos of the remarks made – especially in §17 and §67 – in opposition to Wittgenstein's former dogmatic standpoint, Professor Schlick now informs me that for some years, in writings as yet unpublished, Wittgenstein has taken the view that the rules of language can be chosen freely. Perhaps his view too is developing in the direction of the Principle of Tolerance. (UEBEL, 2009, p. 59, quoting from unpublished letters from Carnap to Schlick).

After reading this, Schlick wrote another letter to Carnap reinforcing the similarities between Wittgenstein's ideas and the Principle of Tolerance. It is clear from Carnap's response that his ideas are different from Wittgeinstein's:

> I myself do not have the impression that Wittgenstein adopts the conception which I designate as the Principle of Tolerance. To be sure, it seems as if he now adopts a more tolerant conception than he (and we all) adopted earlier on. But according to what I have learnt from you (especially from the last paper) and from Waismann, his views do not coincide wholly with

> mine on this point. (E.g., he rejects, if I am informed correctly, sentences that cannot be conclusively verified; moreover, you, and so I suspect he as well, allow as analytic sentences (tautologies) only those for which we possess a decision procedure.)We can talk about these questions later on at our leisure. Here what matters is only that I do not believe that we are in agreement. (UEBEL, 2009, p. 60, quoting from unpublished letters from Carnap to Schlick).

Therefore, as important as Wittgenstein has been to the Vienna Circle, this relevance is limited, in general terms, to the aspects from the *Tractatus* that could be incorporated "as far as we could assimilate them to our basic conceptions" (CARNAP, 1963, p. 24–5). Specially, the acceptance of freedom of choice regarding languages did not have any impact on the development of the Principle of Tolerance.

Logical pluralism

We reach now the final section of this paper. We will investigate whether the Principle of Tolerance implies a logical pluralism and, if so, of which kind.

In general terms, logical pluralism is the claim that there is more than one adequate, coherent, or even, true logic. Note that, from a pure abstract point of view, this thesis may seem trivial today. It's obvious that there are different pure logics. But logical pluralism is not exactly, or not only, about this. It amounts to acknowledge, for instance, that given a certain domain, there are at least two logics that formalize it in a fundamentally different way. And that, nonetheless, both are equally adequate for this task. There are many kinds of logical pluralism (Cf. BEALL, RESTAL, 2006 and SHAPIRO, 2014), but that's not really relevant for the moment. In a schematic way, a pluralist logician claims that exists situations such that

i) β is a logical consequence-in-L from α and $\sim\alpha$; and

ii) β is not a logical consequence-in-L* from α and $\sim\alpha$.[6]

So, there are at least two distinct logics that evaluate differently the validity of the same argument (BEALL, RESTALL, 2006; RESTALL, 2001). From what was discussed in the previous sections we can affirm that the Principle of Tolerance does not imply necessarily a logical pluralism. For, in the first place, it does not claim that every linguistic framework is legitimate. Analyzing the following passage, it is clear the Principle of Tolerance has its limits:

> According to my principle of tolerance, I emphasized that, whereas it is important to make distinctions between constructivist and nonconstructivist definitions and proofs, it seems advisable not to prohibit certain forms of procedure but to investigate all practically useful forms. It is true that

[6] Where α and β are metavariables for formulas, $\sim$ is the symbol for negation, and L and L* denotes different logics.

> certain procedures, e.g., those admitted by constructivism or intuitionism, are safer than others. Therefore it is advisable to apply these procedures as far as possible. However, there are other forms and methods which, though less safe because we do not have a proof of their consistency, appear to be practically indispensable for physics. In such a case there seems to be no good reason for prohibiting these procedures so long as no contradictions have been found. (CARNAP, 1963, p. 49).

Note that we don't have so much freedom to formulate a language. The Principle of Tolerance does not embrace contradictory languages. And the reason is that, despite his tolerance regarding different languages, Carnap still endorses the now called principle of explosion that states that from contradictories premises it is possible to deduce any conclusion. This is clear from the following quote, on which Carnap presents in an informal way the distinction between analytical and synthetic sentences:

> an analytic sentence is absolutely true whatever the empirical facts may be. Hence, it does not state anything about facts. On the other hand, a contradictory sentence states too much to be capable of being true; for from a contradictory sentence each fact as well as its opposite can be deduced. A synthetic sentence is sometimes true – namely, when certain facts exist – and sometimes false; hence it says something as to what facts exists. *Synthetic sentences* are the *genuine statements about reality*. (CARNAP, 1937, p. 41).

That is why, even allowing the formulation of different languages, a proof of the consistency of a given language is still used as a security parameter. Besides that, Carnap never claims that two distinct languages are equally legitimate, which is a core statement of logical pluralism.

The Principle of Tolerance limits itself to allow for different languages to be presented, and for its consequences to be evaluated according to purely pragmatic criteria. Yet, such tolerance opens the possibility for a logical pluralism. But what kind of pluralism is that?

To answer this question, it is enlightening to recall a commentary by Quine – one of Carnap's students - which is well known for those who study non-classical logics:

> whoever denies the law of excluded middle changes de subject. This is not to say that he is wrong in so doing. In repudiating "p or ¬p" he is indeed giving up classical negation [...]; and he may have his reasons." (QUINE, 1960, p. 100).

To put in other words, when a classical logician claims that a given proposition is a logical law, and another logician claims the opposite, they are talking past each other, that is, they are talking about different things.

This reasoning is already present in Carnap's thought. By claiming that each one is free to choose a logic, Carnap gives a huge step towards a logical

conventionalism and pluralism. Nonetheless, since accepting a logic implies in accepting a language that formalizes it, we have at the end a pluralism of languages. This means that, in particular, a classical and an intuitionist logician are not discussing the same subject; they are talking in different languages and, therefore, are even talking about different mathematics. In the final analysis, there's no disagreement between a classical and an intuitionist logician, just a difference in their languages. There's no difference in the way they evaluate a single argument, but rather in the very form of the argument.

Hence, if we compare to the original scheme presented earlier, Carnap's logical pluralism asserts that

i) β is a logical consequence from α and the L-negation of α; and

ii) β is not a logical consequence from α and L*-negation of α.

But that's not all. From Carnap's point of view, the previous formulation of logical pluralism is just incoherent. Each logic comes with a subjacent language. Therefore, once this language is set forward, there's no internal dispute regarding the validity of a given argument. This means, using today's terminology, that Carnap admits the thesis that changes in the syntactical rules of a logical connective implies changes in its meaning.

Hence, the analytical character of each logic is preserved, as well as their universality. Nonetheless, each logic is universal only in its own domain; there is no possibility of interaction between them.

Final Remarks

In sum, even though it is necessary to recognize the historical value of the principle of tolerance and the logical pluralism presented by Carnap, today, this form of pluralism is too restrictive for someone trying to defend the idea that there are more than one adequate logic. For, in the final analysis, Carnap proposes that we develop different logics, and decided about its usefulness, in a sense, according to the purpose of this logic. At no time there's an explicit defense of the possibility of two logics being equally adequate. And, even though it's not possible to talk about adequacy outside a logical system, Carnap still believes there are criteria to determine the usefulness of a logic, such as safety, or simplicity. In the same way, Carnap rejects the possibility of a language containing contradictory sentences, which indicates that his principle of tolerance is not that tolerant after all.

Hence, if someone wants to defend a logical pluralism, Carnapian pluralism is not a good choice. Firstly, because it does not explicitly argue for the adequacy of rival logics, but merely states the possibility of formulate them and investigate their consequences. Secondly, it amounts to a form of pluralism with respect to languages, that is, different languages disagree because they are speaking about different things. And thirdly, for a pluralism

within the same language is not only rejected, but prohibited as a starting point: its formulation is nonsense, and this is based on the assumption that the meaning of a logical constant is given by its syntactical rules.

Bibliography

AWODEY, S.; CARUS, A.W. Carnap's dream: Gödel,Wittgenstein, and Logical Syntax. *Synthese*, v. 1, 159, 2007. p. 23-45.

_____. From Wittgenstein's Prison to the Boundless Ocean: Carnap's Dream of Logical Syntax. In: WAGNER, P. (Ed.). *Carnap's Logical Syntax of Language*. London: Palgrave Macmillan, 2009.

BEALL, JC.; RESTALL, G. *Logical pluralism*. New York: Oxford University Press, 2006.

CARNAP. *Intellectual Autobiography*. In: SCHILPP, P. A. (Ed.). *The Philosophy of Rudolf Carnap*. Illinois: Open Court Publishing Company, *1963*.

_____. *Formal and Factual Sciences*. In: FEIGL, H.; BRODBECK, M. (Eds.). *Readings of philosophy of science*. New York: APPLETON-CENTURY-CROFTS, 1953.

_____. *Foundations of Logic and Mathematics*. In: NEURATH, O. (Ed.). *International Encyclopedia of Unified Science*. Chigado: The University of Chicago Press, 1939.

_____. *The Logical Syntax of Language*. London: Kegan Paul, 1937.
FRIEDMAN, M. *Reconsidering Logical Positivism*. New York: Cambridge University Press, 1999.

POINCARÉ, H. *Science and Hypothesis*. London: Scott, 1905.

QUINE. *Word and Object.* Cambridge: MIT Press, 1960.

RESTALL, G. Carnap's Tolerance, Language Change and Logical Pluralism. *Journal of Philosophy*, v. 99, 2002. p. 426-443.

RICKETTS, T. Carnap's Principle of Tolerance, Empiricism, and Conventionalism. In: CLARK, P; HALE, B. (Eds.). *Reading Putnam*. Oxford: Blackwell, 1994.

SHAPIRO, S. *Varieties of logic*. New York: Oxford University Press, 2014.

UEBEL, T. *Carnap's Logical Syntax in the Context of the Vienna Circle*. In: WAGNER, P. (Ed.). *Carnap's Logical Syntax of Language*. London: Palgrave Macmillan, 2009.

ARGUMENTOS
Revista de Filosofia

Lucas Baccarat Silva Negrão de Campos*

Neurath on context of discovery *vs* context of justification

RESUMO

A distinção entre contexto de justificação e contexto de descoberta é tida como uma das características mais marcantes do empirismo lógico. Nesse sentido, o pressupostos de que todos os empiristas lógicos concordavam quanto à validade irrestrita da distinção é amplamente sustentado na historiografia da filosofia. Ao nosso ver, contudo, esta pressuposição não é completamente correta. Como nós pretendemos demonstrar Otto Neurath, inquestionavelmente identificado como empirista lógico, não concordaria com diversas das formulações da distinção. Por fim nós procuramos demostrar, seguindo a sugestão de Thomas Uebel, que muito embora Neurath rejeite versões mais estritas da distinção, ele não ofereceria objeções a algumas reformulações contemporâneas da mesma.

Palavras-chave: Neurath; Empirismo Lógico; Contexto de Descoberta; Contexto de Justificação; Hoyningen-Huene.

ABSTRACT

The distinction between context of justification and context of discovery is held by many as one of the most distinct characteristics of logical empiricism. In that sense, the presupposition that all logical empiricists agreed on the unrestricted validity of the distinction is broadly sustained in the historiography of philosophy. In our opinion, however, this presupposition is not entirely correct. As we intend to show Otto Neurath, unquestionably identified as a logical empiricist, would not agree with many of the formulations of the distinction. Lastly, following Thomas Uebel suggestion, we try to show that, even though Neurath rejects the strict version of the distinction, he would not object to some contemporaries reformulations of it.

Keywords: Neurath; Logical Empiricism; Context of Discovery; Context of Justification; Hoyningen-Huene.

* Universidade de São Paulo (USP). lucas.baccarat@gmail.com

Introduction

The distinction between context of discovery and context of justification is, without any doubts, one of the most relevant and controversial themes in the recent history of philosophy of science. Especially in the second half of the XX century, under the light of the works of Hanson and Kuhn, the content and extension of the so-called *contexts distinction* went through constant criticism and revision. As a result of the intense dispute surrounding it, whose limits are far from being well delimited, the usefulness of the distinction was questioned or even throughout rejected.

Among those who criticize the context distinction, one can identify, even though only in a very vague sense, two main attitudes towards its adequacy. On one hand, there are those, briefly mentioned above, who completely reject the distinction, arguing that it would promote unnecessary and even dangerous restrictions upon the philosophical analysis of sciences. This alternative, which is often associated with the *strong program in the sociology of science* (BARNES, 1972; BLOOR, 1991), is especially popular among practitioners of a historically and sociologically informed philosophy of science. On the other hand, there are those who recognize that the criticism directed towards the context distinction draws attention to the need of revising it, but sustain that it is not totally useless (HOYNINGEN HUENE, 1987, 2006; NICKELS, 1980; STURM and GIGERENZER, 2006). For the supporters of this second alternative, therefore, one should not simply discard the *contexts distinction* - what is actually needed is a global reappraisal of it, trying to formulate a less strict version of the distinction that would gather only its essential aspects. A very interesting feature of some researchers engaged in the development of an new and refined *contexts distinction* concerns the attempt to clarify the debate, which is frequently very confuse, detailing the various versions of the distinction that were put forward during the dispute.

It is curious, however, that both of the camps just described consent in regarding logical empiricism as the main source of the problems related to the distinction. Although its historical origin is often disputed, it seems to be a consensus among the current participants of the debate that Reichenbach was the first to clearly state and defend the most strict and problematic version of the distinction, which, later on, would be held by every logical empiricist. In fact, the distinction is often regarded as an expression of the logical empiricist program of reducing philosophy to the logical analysis of language or, to put in precise terms, scientific claims. The strict version of the *contexts distinction* is taken to be so bounded to logical empiricism that the downfall of this philosophical movement in the 60's and 70's, would mark the beginning of the questioning of the distinction.

The linking of logical empiricism to the *contexts distinction* is not entirely improper, as most of its members did actually assume the unrestricted validity

of the strict version of the distinction. However, that is not the whole story. In fact we think that the generalization, which states that every member of logical empiricism was committed to the strict separation of discovery and justification, is based in a false presupposition, which nevertheless is very common in the standard historiography of philosophy. According to this presupposition, logical empiricism can be seen as a homogenous group of philosophers, whose disagreements were negligible if compared to the overall agreement regarding the fundamental issues of philosophy of science and epistemology.

As the current work of the recent scholarship of logical empiricism shows (STADLER, 2001; UEBEL 2007), although its is possible to verify a set of shared assumptions among its participants, most logical empiricists had deep and important philosophical divergences, especially if we take the Vienna Circle into account. In this scenario, the strict contexts distinction seems to have also been a controversial topic. In order to prove this last statement, from now on we will focus on the work of the former Vienna Circle member Otto Neurath, who, in our opinion, cannot be looked on as an adherent of the strict version of the *contexts distinction*[1], especially if we understand it as a demarcation criterion between philosophy of science on one hand, and history, sociology and psychology of science on the other hand. As far as we can see, Neurath advocates a sociologically and historically informed philosophy of science and acknowledge the relevancy of empirical research (history, sociology and psychology) in the justification of the decision between empirically equivalent theories and, in a more radical sense, in the acceptance of observational propositions or protocol sentences. Moreover, we argue that even though Neurath does not agree with the strong version of the *contexts distinct*ion, he would not goes as far as completely denying its utility, such that his thinking is actually compatible with some of the contemporary reformulations of it, especially with the *lean distinction* proposed by Hoyningen-Huene.[2]

However, before we engage in the analysis of Neurath's arguments, we would like to give a more clear account of the content of the distinction between context of discovery and context of justification.

Context of discovery *vs* context of justification

In the current work we will heavily rely on Hoyningen-Huene's (HOYNINGEN-HUENE 1987, 2006) presentation of the quarrel surrounding

[1] Howard, 2006 convincingly argues that Reichenbach's main target, in stating the *contexts distinction* was Neurath.

[2] This interpretation is based on Thomas Uebel's appraisal of the topic. In our work, however, we give an more detailed explation of Heunigens-Huene criteria and its relation to Neurath.

the *contexts distinction*. According to him, the first step one should take in approaching the topic is showing how ambiguous the distinction might be and the various forms in which it appears in the multiple texts that address its correctness. According to him, one can recognize five different versions of the *contexts distinction*:

1) The contexts distinction is a distinction between two different processes: the process of discovery and the process of justification. The main point here is that those processes would be temporally distinct, such that the discovery process would precede the justification process.
2) The *contexts distinction* separates the process of discovery on one side from methods of justification on the other side. The opposition here is between factual historical processes and methods, regardless how vague it sounds.
3) The distinction of contexts emphasizes the strictly empirical character of discovery on one hand and the strictly logical character of justification on the other hand.
4) The distinction would demarcate the limits between the domain of research of philosophy of science and that of history, sociology and psychology of science.
5) The contexts distinction is essentially a distinction between the perspectives according to which we pose questions about scientific claims and theories. In that sense, in the context of discovery we might ask: For any given p, how did someone come to accept p? In the context of justification, in turn, the proper question would be: Is p justified?

Once Hoyningen-Huene identifies the various ways in which the *contexts distinction* might occur, he then argues that the commonly reject distinction (the strict one), which is the one associated with logical empiricism, is the one that results from the combination of the versions 1 to 4 above. In our work we will assume this characterization of the logical empiricist conception of the contexts distinction as paradigmatic and try to show that it cannot be applied to Neurath. However, instead of analyzing how Neurath would relate to each one of the versions presented above, we will focus in showing that Neuraths philosophy is actually incompatible with some hidden assumptions about justification that derives from the conflation of versions 1 to 4. As Hoyningen-Huene says, the combination of versions of the contexts distinction implies that the only methods of justification are the logical ones, which, in that sense, would also be the only ones of philosophical interest. If we relocate this claim to the context of the debates that took place in the Vienna Circle, the distinction implies that the only task of philosophy of science is the logical examination of the relations

between the protocol sentences and the theoretical statements, that is, the justification would be restricted to internal issues.

Let us now return to Neurath.

Neurath and the contexts distinction

First of all, we must remark that, even though we sustain that Neurath disagreed with the strict version of the *contexts distinction*, he has never explicitly addressed the topic. Thus, the task we set ourselves to accomplish is not an exposition of the Neurath's actual refutation of the distinction, but an attempt to reconstruct his possible arguments against it. In our opinion, in Neurath's writings, one can easily see that he wasn't in agreement with the strict version of the distinction, principally if one takes note of his description of theory choice and of the pragmatic conditions of the acceptance of protocol sentences.

Neurath constantly addresses the problem of theory choice. From his early writings (which displays a striking continuity with his mature philosophy), the Austrian philosopher continuously stress the need of choosing one among multiple empirically equivalent theories, when there is no logical way of determining the best one. In 1913, for instance, his argumentation runs as follows: Quoting the *Discourse on Method* Neurath says that Descartes was very much right in stressing the need to assume a set of provisional rules for practical purposes, given that from time to time one must choose between equivalents courses of action and, therefore, must act under insufficient insight. Regarding theoretical investigations, however, Neurath's opinion of Descartes is no longer so approving. According to him, Descartes is mistaken in assuming an in principle distinction between theory and practice, where there is only but a degree differentiation. This mistake leads the French philosopher to dismiss the set of provisional rules for theoretical endings, implying that theoretical questions should only be answered when one is in possession of complete insight:

> It was a fundamental error of Descartes that he believed that only in the practical field could he not dispense with provisional rules. Thinking, too, needs preliminary rules in more than one respect. The limited span of life already urges us ahead. The wish that in a foreseeable time the picture of the world could be rounded off makes provisional rules a necessity. But there are fundamental objections to the Cartesian view. Whoever wants to create a world-view or a scientific system must operate with doubtful premises. Each attempt to create a world-picture by starting from a tabula rasa and making a series of statements which are recognized as definitively true, is necessarily full of trickeries. The phenomena that we encounter are so much interconnected that a one-dimensional chain of statements cannot describe them. The correctness of each statement is related to that of all the others. It is absolutely impossible to formulate a single statement about the world without making tacit use at the same

> time of countless others. Also we cannot express any statement without applying all of our preceding concept formation. On the one hand we must state the connection of each statement dealing with the world with all the other statements that deal with it, and on the other hand we must state the connection of each train of thought with all our earlier trains of thought. We can vary the world of concepts present in us, but we cannot discard it. Each attempt to renew it from the bottom up is by its very nature a child of the concepts at hand. (NEURATH, 1983, p. 3).

As it is clear in the passage just quoted, contrary to Descartes, Neurath understands that thinking too necessarily makes use of provisional rules, which should guide the decision between equivalent theories. Later on the text, Neurath dubs those rules auxiliary motives, which, in fact, are motives that don't add up anything new to the question in terms of content, but that, nevertheless, helps the hesitant person. Underlying this reasoning is Neurath's radical antifoundantionalism and his acceptance of the Duhem's[3] holism and underdetermination thesis[4].

However, there are more elements involved in Neurath's description of theory choice then the ones just mentioned. Besides the fact that he recognizes the necessity of making choices in science and the unavoidable need to operate with doubtful premises, Neurath's philosophy is also marked by the strong conviction of the *"irreducible contextuality of knowledge and justification"* (UEBEL, 2007, p.98). Contrary to the standard view on the Vienna Circle, Neurath has never doubted the existence of historical and sociological determinants of knowledge. In his Vienna Circle Days, he would loudly say that "our thinking is a tool, it depends on historical and social conditions [...] we owe our means of expression, our rich language and script". (NEURATH, 1983, p. 46).

If we now gather together the multiple features that integrate the neurathian description of theory choice in science, we have the following situation. On one hand, given holism and the underdetermination of theory by data, choices will always be needed in science. On the other hand, scientific knowledge, just like knowledge in general, does not enjoy any kind of social neutrality, *i..e,* it is also subsumed to historical and social conditions. Now, in order prove that Neurath rejected the strict version of the *contexts distinction*, we must also show that he would allow for sociological explanation of the acceptance or validation of scientific theories, such that the context of justification would have to cover more then just logical methods. But that seems to be precisely the case here:

[3] Duhem's was a major influence on Neurath, who got into contact with the French conventionalists during his participation with Hahn and Frank on the so-called first Vienna Circle

[4] In general terms, the underdetermination thesis states that theory is logically underdetermined by data, since for a given set of data whatsoever, there will always be more then one theory that can account for it.

> Correct thinkers find that, besides the unscientific, the metaphysical, the normative and other ways of considering sociological matters there are also strictly scientific ones that may differ amongst themselves! But this applies also to the physicist who shares the same standpoint. It is conceivable for differences to emerge amongst scientific sociologists that turn on assumptions which one theorist considers just about acceptable whereas another rejects them! Already due to the insufficiency of our knowledge of the available data our predictions are multiply ambiguous! It is resolution that must decide! And this is often historically determined by traditional forms of cognitive cooperation. (NEURATH, 1981, p. 352)[5]

As Thomas UEBEL (Uebel, 2000, p. 144) rightly notices, Neurath here states loud and clear that the decisions between empirically equivalent scientific theories are frequently determined by historical and sociological factors, and, therefore, opens room for sociological and historical explanations of the validations and/or acceptance of scientific theories, that is, Neurath stresses the possibility of external influences to be relevant in the justification of theory choice[6].

The point just made gets even stronger and interesting when we take into account that for Neurath the extension of the domain of underdetermination[7] covers highly abstract scientific theories as well as the protocol sentences[8], that is, according to Neurath even the most elementary statements of system of science are subject to being revised. In that sense, historically determined choices in science are often responsible not only for the selection between empirically equivalent theories, but also for the determination of the set of statements that composes the empirical basis of science.

Given all the arguments presented, we believe it is clear that Neurath would have rejected the strict version of the contexts distinctions, since he acknowledges other elements, besides the logical ones, as being important for the justification of theory choice. For him sociology and history of science and even cultural and political values are held to be valid means of justification of scientific claims. We now ask if the neurathian rejection of the strict version

[5] The English translation of Neurath's original quoted above was extracted from Uebel, 2000, p.144.

[6] As Howard 2006 correctly remarks, Neurath here allows for values to play a significant role on the determination of which theory prevails.

[7] We take the expression "domain of undertermination" from Don Howard (HOWARD, 2003, p 43 and HOWARD, 2006, p. 10). According to him it designates the ambit of application of the underdetermination thesis after logic and experience are allowed to do their work.

[8] About the possibility of revision of protocol sentences Neurath says: "There is no way to establish fully secured, neat protocol statements as starting points of the sciences. There is no tabula rasa. We are like sailors who have to rebuild their ship on the open sea, without ever being able to dismantle it in drydock and reconstruct it from the best components. Only metaphysics can disappear without trace. Imprecise 'verbal clusters' ['Ballungen'] are somehow always part of the ship. If imprecision is diminished at one place, it may well re-appear at another place to a stronger degree." (NEURATH, 1983, p. 92). For a consistent and convincing appreciation of Neurath's protocols (Cf. UEBEL 2007, chapter 11).

of the contexts distinctions implies the rejection of every other formulation of it. As we have already said, this does not seem right and Neurath's stand is compatible with a weaker version of the distinction.

Neurath and the lean distinction between context of discovery and context of justification

A version that is, in our opinion, compatible with the neurathian thinking is the one proposed by Hoyningen-Huene, which is called the *lean distinction*. This version includes both versions 2 and 5 presented above. The core of this distinction is the opposition between a factual and descriptive ambit of investigation on one hand, and an essentially normative and evaluative ambit of investigation on the other. According to this *lean* version, in the context of discovery we are concerned with facts and their description, what would also include the description of epistemic claims. The context of justification, in turn, refers to evaluation of singular claims in accordance with epistemic norms. The version is called *lean,* because it does not imply a demarcation criterion or a distinction between two temporally distinct contexts. Moreover, the simple distinction between the factual and the normative does not imply in any assumption regarding the nature of the facts described or of the epistemic norms.

In our opinion, Neurath would not object to this kind of formulation of the contexts distinction. The fact that he acknowledges non logical procedures as valid methods of justification does not mean that he rejects logical explanatory means of justification in the philosophy of science. Actually most of his Vienna Circle writings stress the benefits science gets form logical clarification of the scientific language and investigations of the logical relations between protocol sentences and more abstract statements. Neurath has never gone as far as denying the possibility of normative theories of epistemic justification. As far as we can see, he only argued that sociological and historical investigations in science could, in fact, inform norms of science.

Regarding his acceptance of the underdetermination thesis. The fact that Neurath recognize that under the duhemian thesis one can see external empirical explanations as contributing for justification issues, does not lead him to advocate that one would be, therefore, obliged see every justificatory explanation as external in character, such that the contexts overlaps. All that Neurath does is arguing in favor of the enlargement of the context of justification, in order to allow sociology and history in. In this sense, given that the *lean distinction* does not say anything about epistemic norms, nothing would prevent historical and sociological claims of being utilized as such. The same goes for the context of discovery, in which validation and acceptance of epistemic claims can be regarded as a historical fact.

This last emphasized feature of the *lean* distinction, it seems us, stand a good chance of capturing Neurath's thoughts on the topic. As long as we allow historical and sociological informed epistemic norms to play a part in the context of justification, he would never object to the possibility of distinguishing a normative domain of inquire and a descriptive domain of inquiry.

References

BARNES, B. Sociological Explanation and Natural Science: A Kuhnian Reappraisal. *Archives Européens de Sociologie* v. 13, 1972, p. 373–393.

BLOOR, D. *Knowledge and Social Imagery,* 2nd ed. Chicago: University Press 1991.

HOWARD, D. Two Left Turns Make a Right: On the Curious Political Career of North American Philosophy of Science at Mid-century. In: RICHARDSON, A and HARDCASTLE, G. (Eds.). *Logical empiricism in North America*. Minneapolis: University of Minnesota Press, 2003 p. 25–93.

_____. Lost Wanderer in the Forest of Knowledge: Thoughts in the Discovery-Justification Distinction. In: SCHICKORE, J. and STEINLE, F. *Revisting Discovery and Justification*: historical and philosophical perspective on the context distinction. Springer, 2006.

HOYNINGEN-HUENE, P. Context of Discovery and Context of Justification *Studies in History and Philosophy of Science.* v. 18, 1987, p. 501–515.

_____. Context of Discovery versus Context of Justification and Thomas Kuhn. In: SCHICKORE, J. and STEINLE, F. *Revisting Discovery and Justification*: historical and philosophical perspective on the context distinction. Springer, 2006.

NEURATH, O. *Empiricism and sociology.* Edited and translated by Marie Neurath and Robert Cohen. Dordrecht: Reidel, 1973.

_____. *Gesammelte philosophische und methodologische Schriften.* Vols. I-II. Edited by R. Haller and H. Rutte. Vienna: Holder-Pichler-Tempsky, 1981.

_____. *Philosophical Papers 1913-1946.* Edited and translated by R. S. Cohen and M. Neurath. Dordrecht: Reidel, 1983.

OKASHA, S. The Underdetermination of Theory by Data and the "Strong Programme" in the sociology of science. *International Studies in the Philosophy of Science*. London, v. 13, n. 3, 2000.

STURM, T. and GIGERENZER, G. How can we use the Distinction Between Discovery and Justification? On the Weakness of the Strong Programme in the Sociology of Science. In: SCHICKORE, J. and STEINLE, F. *Revisting Discovery and Justification*: historical and philosophical perspective on the context distinction. Springer, 2006.

UEBEL, T. Logical Empiricism and the Sociology of Science: The Case of Neurath and Frank. *Philosophy of Science*, Chicago, v. 67, 2000, p. 138-150.

_____. *Empiricism at the Crossroads:* The Viena Circle's Protocol-Sentence Debate. Chicago: Open Court, 2007.

ARGUMENTOS
Revista de Filosofia

Lucas Alessandro Duarte Amaral*

Remarks on the theoretical context of Cassirer's philosophical project

ABSTRACT

In this paper we aim to expose and to analyze some important features in the context of Cassirer's epistemology in his first major work *Substanzbegriff und Funktionsbegriff* (1910), specifically on the problematic relationship between philosophy and science in 19th century. To fulfill our task we opt to proceed in this way: we shall start announcing the problem faced in this period; then we pass to treat the philosophical heritages; in a third moment we shall deal with the scientific legacies, and finally we shall conclude the article with some remarks on the importance of the two referred moments to the origins of Cassirer's philosophical project.

Keywords: Cassirer; Philosophy; Science; 19th Century.

RESUMO

Neste artigo propomo-nos expor e analisar alguns importantes aspectos do contexto epistemológico de Cassirer em sua primeira grande obra *Substanzbegriff und Funktionsbegriff* (1910), especificamente acerca da problemática relação da filosofia com a ciência no século XIX. A fim de cumprirmos nossa tarefa, optamos por proceder desta maneira: iniciaremos anunciando o problema enfrentado nesse período; a partir daí passaremos a tratar as heranças filosóficas; em um terceiro momento trabalharemos os legados científicos e, finalmente, concluiremos o artigo com algumas considerações sobre a importância dos dois momentos referidos às origens do projeto filosófico cassireriano.

Palavras-chave: Cassirer; Filosofia; Ciência; Século XIX.

* Bachelor's degree in Philosophy – PUC-SP (2010); M.A. in Philosophy (CAPES) – PUC-SP (2013); Ph.D. student in philosophy (CAPES) – PUC-SP. Brazil. E-mail: lucasalessandro@hotmail.com; lucasadamaral@gmail.com

Introduction

The philosophy of Ernst Cassirer (1874-1945) represented the culmination of the movement of the Neo-Kantianism of Marburg.[1] From this assumption, we have that the theoretical influences of Cassirer's doctrine came from two different ways: the first one comes, as an immediate result, from the proper development of the Neokantianism of Marburg – in which we have to highlight the thoughts of Hermann Cohen (1842-1918) and Paul Natorp (1854-1924) – and the second one reassembles a broad theoretical context in which other two fundamental points stand out. From one part, it is a philosophical moment and, from another part, it is a scientific moment. Thus, if this large context of debate between philosophy and science in the 19th century is presupposed by the predecessors of Cassirer in Marburg – as well by the Neo-Kantian movement in general[2] – and to the philosopher himself, our task here is to expose it and evaluate it. We will see in the end that these remarks will be of great importance to Cassirer and particularly regarding the importance assumed by the natural sciences[3] on one of his first work the already mentioned *Substanzbegriff und Funktionsbegriff*.

The state of art

The relationship between philosophy and science in 19th century is complicated, if we want to say the minimum. On the one hand, the philosophical hegemony of Hegel seemed consolidated, and, from another, the successful results of science, viewed as an autonomous field of knowledge, were undeniable. Moreover, the distinction between the *Naturwissenschaften* and the *Geisteswissenschften* – such distinction developed by W. Dilthey (1833-

[1] In German there is at least three important schools of Neo-Kantianism which we resume here: (i) The Marburg School (with: Cohen, Natorp and Cassirer); (ii) The Baden School (with: Windelband, Rickert and Lask); (iii) The Realistic School (with: A. Riehl).

[2] Even though all Neo-Kantians had as their background this context, we know that there is a huge difference between the Neo-Kantian schools and also between the members of the current. Since Neo-Kantianism it is a multifaceted movement sometimes certain issues which are questioned by a certain author, are not even mentioned by others. To give a concrete example of a high-importance author in neo-Kantianism, let us take into account Windelband (1848-1915) and the problem of method. Notably, this was one of the well-crafted themes in the doctrine of the Badenian Philosopher. Windelband proposed in his project, roughly speaking, that the role of philosophy would be to evaluate the methods of science, not merely in the sense of research technique, but as a discipline that investigates the conditions of possibility of production of scientific knowledge. In other words, the philosophy evaluates what is established by the science as a starting point, namely, the facts (in the empirical sciences) and axioms (in the formal sciences). See, for instance Windelband's book *Die Prinzipien der Logik*. *Encyclopädie der philosophischen Wissenschaften* (1913).

[3] Obviously, the contribution made by Cassirer in *Substanzbegriff und Funktionsbegriff* is not limited solely to analyze the natural sciences, but also about the formal sciences (e.g., logic) as well as the methods of these sciences. In the end of this article we will mention something regarding this subject.

1911)[4] and resumed by Cassirer himself in his *Essay on Man* (1944)[5] – puts philosophy in a delicate position. If for a long time Philosophy had reach the *status* of the most fundamental discipline of all, through the emancipation of the particular disciplines[6] from its jurisdiction – and let us remember that from that time those same disciplines are possessing their own research methods and objects – what still remains for philosophy? In this sense, one of the day's tasks to be accomplished at that time will be precisely this one: to restore the positive relationship between philosophy and science. Faced with these huge problems, the Neo-Kantian movement would emerge and would accept this difficult challenge of restoring the dialogue between philosophy and science.

Also in regarding to this, take into account that Cassirer, in the first volume of his *Philosophy of symbolic forms* (on language), notes this problem concerning the applicability of the important results achieved in the field of natural sciences, worked by him in his book *Substance and Function* – whose major concern in the field of logic, mathematics and natural science is indubitable – to the field of the *Geisteswissenschften*. Such problem, as already mentioned above, was given by Cassirer's predecessors.[7]

The alternative found by this generation of thinkers will have its starting point signed within the framework of a dialogue on different nuances to return to Kant and the philosophical trends of his time. So much so that it became well known the appeal of Otto Liebmann (1840-1912) of 'return to Kant' on his classic book *Kant und die Epigonen* (1865). In it, at the end of the chapters, Liebmann always concluded with the phrase *'Also muss auf Kant zurückgegangen*

[4] See for instance Dilthey's *Einleitung in die Geisteswissenschaften* (1883).

[5] See specifically on part II, *Men and Culture*, the subjects 'History and Science'.

[6] With the emancipation of particular sciences (economics, social sciences, anthropology, psychology, etc.) from the purview of philosophy, the aspiration of philosophy as a system in which are worked out the various areas of knowledge, it is becoming increasingly a rather complicated task.

[7] See for example Cassirer's first words in his preface of his first volume of *Philosophy of Symbolic Forms* (on Language): "Die Schrift, deren ersten Band ich hier vorlege, geht in ihrem ersten Entwurf auf die Untersuchungen zurück, die in meinem Buche „Substanzbegriff und Funktionsbegriff" (BERLIN, 1910) zusammengefaßt sind. Bei dem Bemühen, das Ergebnis dieser Untersuchungen, die sich im wesentlichen auf die Struktur des mathemalischen und des naturwissenschaftlichen Denkens bezogen, für die Behandlung geisteswissenschaftlicher Probleme fruchtbar zu machen, stellte sich mir immer deutlicher heraus, daß die allgemeine Erkenntnistheorie in ihrer herkömmlichen Auffassung und Begrenzung für eine methodische Grundlegung der Geisteswissenschaften nicht ausreicht. Sollte eine solche Grundlegung gewonnen werden, so schien der Plan dieser Erkenntnistheorie einer prinzipiellen Erweiterung zu bedürfen. Statt lediglich die allgemeinen Voraussetzungen des wissenschaftlichen Erkennens der Welt zu untersuchen, mußte dazu übergegangen werden, die verschiedenen Grundformen des „Verstehens" der Welt bestimmt gegen einander abzugrenzen und jede von ihnen so scharf als möglich in ihrer eigentümlichen Tendenz und ihrer eigentümlichen geistigen Form zu erfassen. Erst wenn eine solche „Formenlehre", des Geistes wenigstens im allgemeinen Umriß feststand, ließ sich hoffen, daß auch für die einzelnen geisteswissenschaftlichen Disziplinen ein klarer methodischer Überblick und ein sicheres Prinzip der Begründung gefunden werden könne. Der Lehre von der naturwissenschaftlichen Begriffsund Urteilsbildung, durch die das „Objekt" der Natur in seinen konstitutiven Grundzügen bestimmt, durch die der „Gegenstand" der Erkenntnis in seiner Bedingtheit durch die Erkenntnisfunktion erfaßt wird, mußte eine analoge Bestimmung für das Gebiet der reinen Subjektivität zur Seite treten." (PSF, I, V)

werden'. Moreover, the contribution made by Liebmann – starting his research with German idealism (Fichte, Schelling, Hegel), following with the realistic aspects (Herbart), and the empiricists aspects (Fries) and concluding with Schopenhauer – suggests that what followed Kant's transcendental philosophy was not something rightly consequential, but on the contrary was a setback. That's why we would have to return to a save harbor (Kant) and the exhortation mentioned above would appear.

The philosophical heritage

The Kantian philosophy at the philosophical context of the end of 18th century until Hegel's death was subject of criticism, in addition to having gone through numerous and the most diverse interpretations. To remind us of some very close to Kant,[8] let us take, *e.g.*, first line names such as: Mendelssohn (1729-1786), Hamann (1730-1788), Jacobi (1743-1819), Maimon (1753-1800) and Reinhold (1757-1823). Subsequent to this first generation of thinkers, a new one would emerge and would be of even greater importance to our present objectives, and this generation, notably, the apex of German idealism, with the exponents of the famous triad: Fichte (1762-1814), Schelling (1775 -1854) and Hegel (1770-1831). Also in regard to German idealism, it is of particularly importance two other points, which we pass to describe below. It is, on the one hand, its relationship with the critical philosophy of Kant and, on the other, its legacy to the Neo-Kantian movement.

The German idealism appears on the philosophical scene of that time critically dialoging with the Kantian philosophy. And that does not mean anything other than idealism emerges as a systematic alternative and more consistent than what the criticism of Kant intended. On such positive consequences of the movement, let us take into account that a number of dualisms, deriving from the old Cartesian scheme, which the author of the *Critique of Pure Reason* had accepted largely in the context of his doctrine, would have been dissolved by idealism.[9] To remind ourselves of a few, let us take these: (i) subject-object; (ii) matter-form; (iii) intuition-concept; (iv) phenomenon-thing-in-itself. Finally, in addition to allegedly dissolved this series of dualisms, idealism had also proposed two important criteria that have become their characteristic marks, namely:

i) totality and

ii) systematicity.

[8] Some of those thinkers had discussed with Kant himself. If we look the exchange of letters of the German Philosopher, then we will see that Mendelssohn, Hamann, Maimon and Reinhold already had spooked with Kant.

[9] See Hegel's *Encyclopedia of Philosophical Sciences* (1817) specially §§ 40-42.

Put in those terms, a philosophy worth its salt should contain, therefore the a character of a system. And an important point to note here with respect to this is the fact that Cassirer will not abandon such an ideal and is considered one of the last, if not the last, philosopher to make a proper system of philosophy, which would be worked various areas of knowledge, such as: science, politics, language, myth, anthropology, etc. In addition, and just as importantly, it is needed to highlight another decisive factor in Cassirer's background, namely, he defends the thesis that all these areas of knowledge mentioned above are equally valid knowledge. In this sense, we would have, for example, that the discourse of science is no more or less important than that of myth. Indeed, within the framework of Cassirer's epistemology in *Philosophy of symbolic forms*, both (science and myth) are equally valid ways of understanding the world.

In order now to clarify some aspects regarding the importance of idealism in German philosophical context of the 19th century, we take into account the particular case. For this, take the example of Hegel – and there is no doubt that this author serves us as a representative model of German idealism.

At first, let us remember that Hegel's system aims to, roughly speaking, a science of absolute, fulfilling in this way with the first criteria mentioned above. And already on this first point the author of the *Phenomenology of Spirit* proposes a change of two central concepts commonly used in philosophy, namely:

1) The very notion of philosophy. Hegel modifies the design of this course, etymologically known as being one discipline which has the 'love of wisdom', to be understood as "the wisdom". In these terms, the philosophy would not be a discipline among many others, but the most important one.
2) His model of science. In this sense, one of the most important implications of this would be that the ideal of science would not be contemplated in Newtonian mechanics, as understood Kant[10] for example, but in the philosophy itself, which would then be considered science[11] *par excellence*.[12]

Another crucial notion to Neo-Kantianism coming of idealism was the spontaneity of the spirit. However, this spontaneity has to be proved precisely

[10] It is well known that Hegel makes a severe criticism of Newton in his *Dissertatio,* 1801. Perhaps one of the most relentless criticisms of Hegel to Newton was that the first accuses the theory of the second to be nothing else then a mere random calculation.

[11] An interesting point regarded to this is that also Husserl would consider the philosophy as "the" science. Just remind us of his famous writing *Philosophy as rigorous science* (1911).

[12] On this subject see Hegel's preface to his *Phenomenology of Spirit*.

in the science itself. In proposing the "transcendental method", Cohen distance to both the Hegelian method (dialectical-speculative) and the psychological method.[13]

At the end, this totalizing aspiration of idealism, besides making possible the loss of track, as reported by Otto Liebmann, eventually leads philosophy to a direct conflict with science. This one, however, follows its profitable course without caring so much about what philosophy has to say about it. Add to this adverse state of things to philosophy the other decisive factor, pointed out once: the emancipation of the particular sciences from philosophy – and both natural sciences and the sciences of spirit came, increasingly strong, claiming its autonomous place in the field of knowledge.

The Scientific heritage

There is no doubt that the emergence of new theories in the field of natural science in the nineteenth century influenced a lot the philosophical theories which intended to speak about science and there are many examples in history that serve as an endorsement of that. To name just a classic example of this, let us remember that Kant, who saw in Newtonian mechanics a model of science, wrote his famous *Critique of Pure Reason* in light of this crucial scientific theory. Like him, other authors had in their particular contexts different scientific theories as models. In this same vein, Cassirer, as we shall see, would work out his doctrine also in view of a model of science. However, it was not in Newtonian mechanics, but the electromagnetism of Maxwell that Cassirer has its well-established science model.

Regarding the importance of certain scientists and their determinant theories, a large list would be made. To remind ourselves of a few examples here, let us take up these names: Mendel, Lamark and Darwin in biology; Weierstrass, Cantor, Dedekind, Boole and Galois in arithmetic; Felix Klein and Gauss in geometry; Kelvin, Boltzmann, Faraday and Maxwell in physics. Some of the direct fruits of these efforts are focused on the following theories: (i) thermodynamics; (ii) non-Euclidean geometries; (iii) logicism of arithmetic; (iv) electromagnetism.

On the importance of Maxwell's Electromagnetism theory to Cassirer's epistemology

If we consider the philosophy of Cassirer, the author who arouses greatest interest is, without doubt, Maxwell and his theory of electromagnetism.

[13] On this subject see: PORTA, M. A. G. *O problema da 'Filosofia das Formas Simbólicas'*. In: *ESTUDOS NEOKANTIANOS*. Loyola, 2011. P. 48-49.

Accordingly, the influence of this scientist is similar to that exerted by Newton in Kant's epistemology.[14]

With the emergence of these new scientific theories of the nineteenth century Newtonian mechanics is called into question. New phenomena come to be studied by scientists of this century and gradually the Newtonian program could not give more account to explain them. What finally mark the fall of Newtonian hegemony in the scientific scene of the time was precisely its inability to interpret what Maxwell proposes with its innovative theory of electromagnetism.[15] While the Newtonian world was a world in which it was possible to intuit, the intuitiveness condition increasingly begins to lose its place in the face of new scientific concepts. Thus, physics has no longer as one of its main tasks to provide a picture of the universe. Furthermore, science has become a discipline in which it investigates the principles instead of a physics seeking to investigate the matter properly. While the concept of material object was considered the fundamental concept of physics at that time this conceptualization changes.

From the new ideas of Faraday and Maxwell, the concept of field comes to occupy a prominent place in physics. The culmination of this radical course in physics is given in Einstein's relativity, having as one of its philosophically relevant points, and essential in the Einstein's program, that relativity is not restricted to the requirements of intuitiveness, culminating thus with the radical break on science whit all intuitive view of the universe. This process results, in a certain sense, from the impossibilities of Newton's mechanics to interpret Maxwell's equations. And Cassirer was a strong proponent of the thesis that thanks to Maxwell, Einstein could do what he did. [16]

Concluding remarks

In the mid-19th century in the *Materialismusstreit* will finally play a decisive role in the roots of neo-Kantianism. Moreover, this controversy will reshapes the neo-Kantian the idealistic worldview. Thus, according to the transcendental method – which opposes both the dialectic-speculative-metaphysical method of Hegel as the psychological method – as mentioned

[14] Another author who has in his philosophical horizon another famous scientific theory is Moritz Schlick, who saw in Einstein's relativity that model. In this sense, we would have: Einstein is to Schlick what Newton and Maxwell were to Kant and Cassirer respectively.

[15] To be fair the mechanical theory of Newton is received in the 19th century as "the" science, but, according to what we said above, this theory would be put into question, among other reasons, by the appearance of new themes and subjects of study in science, besides the notion of field, like for instance the concept of heat. The electromagnetism theory of Maxwell is one of many others theories in science at this moment, as we said above.

[16] See for example in Cassirer's ERT the first chapter: on the concepts of measure ad concepts of things.

above, originally proposed by Herman Cohen, Science will be a *FAKTUM*, that is, a starting point of reflection.

In *Substance and Function*, Cassirer will do a thorough contribution regarding the scientific point briefly exposed here. His analysis will aim to a "logic of objective knowledge"; this project was pointed out firstly by the philosopher in his Article *Kant und die moderne Mathematik* (1907) (See Cassierer's KMM, p. 44). Therefore Cassirer will have to evaluate several points: from the rising of a new logic on the last quarter of the nineteenth century, through the development of arithmetic, geometry, natural sciences, to the methods developed by authors like Mach and Poincaré.

Finally, as we said initially, there remains an important gap to be filled if you want to get to Cassirer's thought. Within this, two other points should be analyzed which relate to two other representative Neo-Kantians of Marburg, who succeeded E. Cassirer: Herman Cohen and Paul Natorp. Notably both were of importance to the philosopher of culture. However, we will leave this task for the next opportunity.

Bibliography

BEISER, F. *After Hegel – German Philosophy 1840-1900.* New Jersey: Princeton University Press, 2014.

CASSIRER, E. *Einstein's Theory of Relativity.* [ERT – 1921]. Open Court. Chicago. 1923.

_____. *El problema Del conocimiento, v. II-III-IV.* [EP, II, III, IV – 1907, 1919, 1922]. Fondo de cultura Econômica. 1953.

_____. *Ensaio sobre o Homem.* [EM – 1945]. São Paulo: Martins Fontes, 2000.

_____. *Filosofia das formas simbólicas. Primeiro Tomo:* A linguagem [PSF, I – 1923] São Paulo: Martins Fontes, 2001.

_____. *Kant und die Moderne Mathematik.* [KMM – 1907] Ernst Cassirers Gesammelte Werke, v. 8. Felx Meiner. Hamburg. 1998.

_____. *Substance and Function.* [SF – 1910]. Chicago. Open Court, 1923.

FRIEDMAN, M. *A Parting of Ways:* Carnap Cassirer and Heidegger. Chicago: Open Court, 2000.

HEIS, J. *Ernst Cassier's Neo-Kantian Philosophy of Geometry.* British Journal for the history of philosophy, 2011.

HEGEL, G. W. F. *Encyclopedia of the Philosophical Sciences in Basic Outline – Part I:* Science of Logic. New York: Cambridge U. Press, 2010.

_____. *Fenomenologia do Espírito.* Petrópolis RJ: Vozes, 2002.

KANT, I. *Critique of Pure Reason.* Tradução Paul Guyer. Cambridge. U. Press,1998.

KÖHNKE, K. *The Rise of neo-Kantianism:* German academic philosophy between idealism and positivism. [1986]. New York: Cambridge University Press, 1991.

PORTA, M. A. G.: ZURÜCK ZU KANT! Adolf Trendelenburg, a superação do idealismo e as origens da filosofia contemporânea. In: PORTA, M. A. G. *Estudos Neokantianos*. São Paulo: Edições Loyola, 2011, p. 15-44.

_____. *DE NEWTON A MAXWELL* – Uma contribuição à compreensão do projeto cassireriano de uma 'filosofia das formas simbólica'. In: PORTA, M. A. G. *Estudos Neokantianos.* São Paulo: Edições Loyola, [s.d.], p. 71-101.

_____. *Cassirer e Kant.* In: PORTA, M. A. G. *Estudos Neokantianos.* São Paulo: Edições Loyola, p.145-184.

RESENDE JR, J. *Em busca de uma teoria do sentido:* Windelband, Rickert, Husserl, Lask e Heidegger. São Paulo: EDUC, 2013.

WINDELBAND, W. *Die Prinzipien der Logik*. *Encyclopädie der philosophischen Wissenschaften* hrsg. A. Ruge. Logik (v. 1). Tübingen, J. C. B. Mohr (p. Siebeck), 1913.

Revista de Filosofia
ARGUMENTOS

Marco Antonio Chabbouh Junior*

On the inefficiency of Lambert's and Mendelssohn's objections against the inaugural dissertation's theory of time**

ABSTRACT

Kant's theory of the ideality of time suffered attacks since it was first conceived in the *Inaugural Dissertation*. Johann Heinrich Lambert and Moses Mendelssohn, two of Kant's most frequent correspondents, were the first to object to that doctrine. In this paper I intend to show that these objections are not successful against the theory of 1770. To achieve that aim, I will firstly explain the structure of the objections, secondly I will show that Kant attacks some epistemological consequences of the postures assumed by these objections and, finally, I will demonstrate how the argument put forward in the first subsection of § 14 of the *Inaugural Dissertation* is the foundation to reject the objectors' assumptions. Additionally, in the last part, I will show that such objections would make sense if the 1770's theory of time was founded on a theory of forms as *temporarily presupposed* in the course of experience. However, I will also show that such an interpretation would transgress both the principle of charity and the literality of certain excerpts of the text.

Keywords: Immanuel Kant; *Inaugural Dissertation;* ideality of time; Johann Heinrich Lambert; Moses Mendelssohn.

RESUMO

A tese kantiana da idealidade do tempo sofreu ataques desde que foi primeiramente concebida na *Dissertação de 1770*. Johann Heinrich Lambert

* Ph.D. student PUCSP/CAPES. Email: m_chabbouh@hotmail.com
** This paper is a small part of a much larger doctoral research on Kant's doctrine of the ideality of time. This research takes place at the Pontifical Catholic University of São Paulo and is supported by CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior).

e Moses Mendelssohn, dois dos mais frequentes correspondentes de Kant, foram os primeiros a objetar contra aquela tese. No presente trabalho eu pretendo mostrar que essas objeções não surtem efeito nem mesmo contra a teoria de 1770. Para isso, primeiro exporei a estrutura das objeções, em seguida mostrarei que Kant ataca textualmente algumas consequências epistemológicas das posturas pressupostas por essas objeções e, por último, demonstrarei como o argumento exposto no primeiro subitem do §14 da *Dissertação de 1770* é o fundamento para contrapor os pressupostos dos objetores. Adicionalmente, na última parte, eu mostrarei que tais objeções fariam sentido se a teoria do tempo de 1770 fosse fundada em uma teoria das formas enquanto *temporalmente pressupostas* no curso da experiência. Contudo, mostrarei também que interpretar de tal maneira viola tanto o princípio de caridade quanto a literalidade de certas porções do texto.

Palavras-chave: Immanuel Kant; *Dissertação de 1770;* idealidade do tempo; Johann Heinrich Lambert; Moses Mendelssohn.

Introduction[1]

In 1770, as a part of his *On the form and principles of the sensible and the intelligible world*, Immanuel Kant develops a theory of time. One of the most controversial doctrines of this theory is that time is ideal and subjective and, therefore, it is not real or objective. To reach this conclusion, Kant makes use of some expositions that clarify the relationship between time, succession and simultaneity and he confronts the results of these expositions with other alternatives to his theory of time.

Also in 1770, Kant receives letters from two of his most frequent correspondents, namely, Johann Heinrich Lambert and Moses Mendelssohn and the two postulate two famous objections to Kant's doctrine of the ideality of time. Lambert objects that there are real changes and that, consequently, time must also be real. Mendelssohn objects that succession is a determination of finite spirits and that those spirits are not only subjects, but also objects represented by other finite spirits and that, therefore, time must determine at least one real thing.

It is a fairly common view among interpreters that Kant does not consider the seriousness of the objections. Two good examples of this attitude are Kemp

[1] As usual, references to Kant's works and correspondence will be to *Kant's Gesammelte Schriften, Akademie Ausgabe* (Ak volume: pages). The only exception will be the use of the standard "A" and "B" in the case of references to Kant's *Critique of Pure Reason*.

Smith (1923, p 122-114) and Kitcher (1993, p. 140-141). Even Paton (1936, p. 182), who maintains the inefficiency of the objections, affirms that Kant did not succeed in understanding it. Falkenstein says that despite the objections being effective against the *Inaugural Dissertion*'s doctrine of time, Kant would have altered his theory in order to overcome the objections at least in the *Critique of Pure Reason* (FALKENSTEIN, 1991, p. 227-228 and 239-240). I, on my part, will argue that in 1770 Kant already had elements not to fall in the apparent inevitability of Lambert's and Mendelssohn's objections.

My first argument in this respect is textual and the second one is a drifting of consequences from the textual argument: firstly I will show that Kant literally expresses his rejection of the leibnizian reductionist/relativist theory of time, theory which is equivalent to the ones maintained by Lambert and Mendelssohn. My second argument consists in pointing out that Kant's justification for such a rejection is epistemological. To reach my aim, I will first present the nature of the objections. Secondly, I will point out the excerpts in which Kant explicitly rejects Lambert's and Mendelssohn's proposals. Thirdly and finally, (i) I will indicate how Kant refutes the epistemological consequences of those proposals and (ii) - against Falkenstein – I will show that the Prussian philosopher did not hold an imposition thesis in the *Inaugural Dissertation*.

Lambert's and Mendelssohn's objections

Kant published the dissertation *On the form and principles of the sensible and the intelligible world*[2] in August 1770. At the time of the publication, he sent a copy to each of his most frequent correspondents. The dissertation reached the hands of the mathematician Lambert and of the philosopher Mendelssohn.

Less than two months later, Lambert sent a letter to Kant in order to express his views with respect to the *Inaugural Dissertation*. A considerable part of the letter's text is addressed to expose Lambert's considerations about Kant's doctrine of time. Lambert says he agrees with every step and with all of the conclusions of Kant's argument on time, except one. The mathematician accepts the thesis according to which time is a necessary condition of sensible apprehension, he accepts the thesis that time is a pure intuition, he regards as true the negative results according to which time is neither a substance nor a relation, but he does not accept the ideality of time (LAMBERT, 1999, p. 106-107).

Lambert offers one argument for the thesis that time cannot be exclusively ideal and the argument has two steps. The first step establishes a connection between time and change - relationship that is usually established by the

[2] From now on simply *Inaugural Dissertation*.

reductionist theories of time[3]. In his words "If changes [*Veränderungen*] are real, then time is real, whatever it may be. If time is unreal, then no change can be real" (LAMBERT, 1999, p. 107). That is, if we find just a single case of actual change then time must also be real. The second step is to show that there is one case of real change. Again in Lambert's words "even an idealist must grant that changes really exist and occur in his representations, for example, in their beginning and ending" (LAMBERT, 1999, p. 107). That is to say, at least in the acts of passing to exist and ceasing to exist of our representations there is change; at least in this case we can't deny that there is real alteration. If there is real change and if there is an inseparable connection between change and time, then time must be real.

Two months later, Kant received another very similar objection. Mendelssohn's criticism also takes a dual path. Firstly, Mendelssohn indicates that "Succession [*Succession*] is after all at least a necessary condition of the representations that finite minds have." (MENDELSSOHN, 1999, p. 110). This means that finite minds – i.e. the subjects - are determined by succession. At this point he seems to be calling attention to the same point already brought up by Lambert: we pass through our representations, we do not merely order them in time. Secondly, he points out that the finite subjects are not merely subjects that represent, but are also objects of representations of other minds. Now the other minds also order their representations in time. Thus, the subject - that is, a real object of the representations of other minds - must be determined temporally and thus time must be something real.

There seems to be a common ground between the two objections. Lambert and Mendelssohn share a premise when they conclude that time is not entirely ideal. This premise is that the subjects do not merely order their representations in time, but they also pass through their representations; they start to have and stop to have this or that representation.

The textual argument

To better understand how Kant develops his theory of time in 1770 is necessary to understand the argumentative way he takes in §14 of the *Inaugural Dissertation*. That section is divided by the philosopher in seven subsections. In the first of these subsections Kant seeks to prove the independence of time relatively to the senses. In the second, he argues that time is a singular representation and therefore intuitive. The third subsection is devoted to summarize the results of the previous two subsections: since time is both an intuitive and pure representation then time must be a pure intuition. The fourth

[3] This is pointed out, for example, by Michael J. Futch (2008, p. 6-7).

subsection aims to prove that time is a continuous magnitude and that it is the principle of the laws of the continuum. The fifth and sixth subsections aim to derive conclusions from the previous expositions: the fifth section derives the negative consequences while the sixth derives the positive consequences. In the seventh subsection Kant summarizes all the exposition situating it in the general plan of the *Inaugural Dissertation* (KANT, 1992, p. 398-402).

To understand in what sense Kant already had the essential elements to answer Lambert and Mendelssohn it is important to consider initially the line of argumentation in the fifth subsection of the §14. After having demonstrated that time is *a priori*, intuitive and a continuous quantity, Kant does a survey of its negative conclusions regarding the nature of time. In that section, he basically takes the results of previous expositions and contrasts with four ontological contemporary alternatives in order to deny them all. These alternatives are (i) that time is either a substance or an accident; (ii) that time is a relation; (iii) that time is a real and existing flux and; (iv) - this is the most important alternative for us here - that time is "something real abstracted from the succession of internal states" (KANT, 1992, p. 401).

The argument to refute the thesis that time would be a substance or an accident is the recognition that in order to coordinate substances and accidents it is necessary simultaneity and succession. However, both simultaneity and succession, says Kant, are only possible by means of the concept of time. Thus, time can be neither a substance nor an accident, because it is a precondition for their coordination.

The second argument is intended to refute the thesis that time is a relation and it follows the same path of the first one. As relations are presented to the senses, these relations have neither a content of succession nor a content of simultaneity. In contrast, relations, insofar as they are presented to the senses, contain only positions which should be determined in time. To determine positions in time is precisely what allows the identification of a successive or simultaneous relation. Therefore, time cannot be a relation, but must be a precondition for the perception of relations.

Kant considers the thesis that time is a real existing and continuous flux which is basically the position taken by Clarke in his *Correspondence* with Leibniz. The problem is that the Prussian philosopher offers no argument against that thesis. He says only that such position is "a most absurd fabrication" (Ak II, p. 401).

Finally, at the end of the subsection 5, Kant offers two arguments against the position that time is "something real abstracted from the succession of internal states" (Ak II, p. 401), a position which he credits to Leibniz. The main argument is that such position incurs in a vicious circle. I have pointed out that, in the refutations of the other theses, Kant considers time as a necessary condition for the apprehension of succession. If this Leibniz's position defines

time as being abstracted from succession, then this position is simply inconsistent with what has been previously demonstrated. Hence, Kant adds in a second argument that the thesis according to which time is something real abstracted from the succession of internal states would cause the movement to determine time and not time to determine the laws of motion – this last point is a kind of prefiguration of what Kant would call "Transcendental Exposition of the concept of time" in the *Critique of Pure Reason* (KANT, 1918, A 32 = B 48-49).

This last position that Kant hopes to have refuted seems to be exactly the position advocated by Lambert and Mendelssohn. According to them, time is something real. Also according to them, the reality of time can be seen at least in the succession which determines finite spirits; it can be seen because of the reality of changes in the rise and in the cease of our representations. Kant's main argument against them, eminently epistemological, would then be the one exposed above: time cannot be conditioned by succession because it is a prerequisite for the perception of succession. If we perceive two events as being successive this is due to the fact that we have a notion of time that conditions that perception.

Succession and apriority of time

Now we should address a second issue. As I said before, Kant's basic argument against the reductionist view of time is that time cannot be abstracted from succession because it is prerequisite for it. However, this cannot be a mere statement; there must be some reason why Kant states that time is independent from successive appearances. Otherwise, Kant would have no way to answer Lambert and Mendelssohn, but worse, he would not even have a way to propose his own theory of time as an alternative theory to Leibniz's.

Kant, as I advanced, offers such an argument, and that argument is the one offered in the first subsection of the §14 of the *Inaugural Dissertation*. In that part of the text, Kant's aim is to demonstrate the precedence of the representation of time relatively to the senses (KANT, 1992, p. 398-399). The proof is achieved by an analysis of our perception. All that we perceive is ordered as successive or simultaneous. Either two objects appear to me as coexisting in the same time span and are, therefore, simultaneous or these two objects appear to me as non-coexisting in the same time span and are, therefore, successive. The only way to perceive these objects as coexistent or as non-coexistent in the same time span is being in possession of a notion of time that must at least be unitary, one-dimensional and progressive. Otherwise, there would be no point in talking about simultaneity or succession since, on the one hand, the lapses would not be part of the same temporal unit and, on the other hand, there would be no way to identify precedence or sequence of two intuitions. Thus, Kant

concludes that for us to perceive something as successive or as simultaneous we must have an independent notion of time and since everything that appears to us is ordered as successive or simultaneous then time does not originate from the senses.

From this it is possible to understand the foundation for the solution of Mendelssohn's and Lambert's objections. Firstly, the perception of succession is conditioned by the notion of time. This applies to the perception of physical objects as well as to the perception of internal states (Ak II, p. 397). Mendelssohn argues that succession is a necessary condition of the representations of finite minds, i.e., that the representations of finite minds succeed each other. Kant does not deny that we are aware of our representations as succeeding each other. However, the Prussian philosopher would argue that we can only perceive our inner states as succeeding each other - and, indeed, any type of succession - because we are in possession of a representation of time that determines the totality of our experience – and, in a very particular way, the experiences of succession and simultaneity.

Falkenstein (1991, p. 228) states that Mendelssohn's objection undermines the *Inaugural Dissertation's* theory of time. He argues that in the 1770's text, Kant held a theory of time according to which we first receive the matter of sensible representations to then apply the form (time and space), but in the *Critique of Pure Reason*, instructed by the objections of his correspondents, the Prussian philosopher would have changed his position. The criticism of Falkenstein would make perfect sense if the 1770's theory of time was in fact a kind of imposition thesis[4]. The biggest problem is that the *Inaugural Dissertation*, even more than the *Critique of Pure Reason*, seems to hold something very distinct. Firstly, in the *Dissertation* Kant explicitly denies the naive innatism which is a possible form of the imposition thesis[5].

> Finally, the question arises for everyone, as though of its own accord, whether each of the two *concepts* [time and space] is *innate* or *acquired*. The latter view, indeed, already seems to have been refuted by what has been demonstrated. **The former view, however, ought not to be that rashly admitted**, for it paves the way for a philosophy of the lazy [...] **But *each of the concepts* [time and space] *has*, without any doubt, *been acquired*** [...] (KANT, 1992, p. 406, **emphasis added**).

Secondly, Kant explicitly states that the form is not completely disconnected from reality as would be in the imposition thesis.

[4] In Falkenstein's definition, according to the imposition thesis "[space and time] are imposed by the mind on the objects of knowledge, as if nothing apart from our mental representations exhibited spatio-temporal properties; rather our minds are so constituted that we inject spatio-temporal form into our mental representations" (FALKENSTEIN, 1991, p. 227).

[5] Namely, the form of imposition thesis sustained by Kemp Smith (1923, p. 89-91).

> Moreover, just as the sensation which constitutes the *matter* of a sensible representation is, indeed, evidence for the presence of something sensible, though in respect of its quality it is dependent upon the nature of the subject in so far as the latter is capable of modification by the object in question, **so also the *form* of the same representation is undoubtedly evidence of a certain reference or relation in what is sensed** [...] (Ibid Ak II: p. 393, **emphasis added**).

Finally and ultimately, the excerpts that led Kemp Smith to defend his version of the imposition thesis are not present in the *Inaugural Dissertation*. Good examples of these excerpts in the *Critique of Pure Reason* would be "its form must lie ready for the sensations *a priori* in the mind" (KANT, 1918, A 20 = B 34) and "Space is represented as an infinite given magnitude" (KANT, A 25 = B 39).

Since there are reasons to argue that Kant's position in 1770 concerning the ideality of time was not a kind of imposition thesis, then there is no reason to affirm the efficiency of Lambert's and Mendelssohn's objections. Such objections point to the fact that there is something in reality that somehow implies the diversity of temporal characteristics in the sensible objects and in particular in the internal intuitions (KANT, 1992, Ak II, p. 393). As I just defended, Kant was ready to accept this since 1770.

Conclusion

It seemed that Kant took the objections of Lambert and Mendelssohn very seriously. Besides having responded them on two different occasions, the philosopher of Königsberg has even claimed that Lambert's objection is "the most serious objection that can be raised against the system" (KANT, 1999, Ak X, p. 134). Therefore one would expect that they have somehow influenced the mature doctrine of the *Critique of Pure Reason*. Nevertheless, I hope to have shown that if there was any influence it was not related to the core of Kant's theory of form and matter.

Both objections derive their strength from the recognition that our representations from the internal sense succeed each other; from the recognition that they change. Such recognition would do great damage if the objections were attacking a theory in which time is a prior representation that at the moment of each perception applies to the matter given and provides it with its temporal features. This is because what Lambert and Mendelsohn are pointing out is that there is something in us and, therefore, in real things, which makes the representation B succeed the representation A and not the other way around; that there is something in a real thing that causes the representation X to be extinguished and the representation Y to be risen.

As I intend to have shown, Kant does not advocate such a theory in the *Inaugural Dissertation*. In addition to the fact that this theory would be inconsistent and in addition to the fact that the excerpts that lead Kemp Smith to defend an imposition thesis are not present in the text of 1770, in the *Inaugural Dissertation* Kant recognizes the acquisition of the notion of time and the relationship of the forms of the sensible world with something real. Thus, to maintain that the objections of Lambert and Mendelssohn made Kant change his theory violates the principle of charity as well as the literality of the text.

Finally, if we interpret the *Dissertation*'s theory of time as an analysis of our experience and accept that time is an acquired notion that, while being the form of the sensible world, should be related to what is felt, then we are able to understand why Kant defends the ideality of time and that such view is not inconsistent with what Lambert and Mendelssohn pointed out in their objections. Time is independent of the succession and simultaneity. All that we perceive, even internally, is subject to time as a form. However, there is something in reality that contributes in some way to the temporal differences of particular events. We actually perceive our representations as succeeding each other; we actually perceive representations as emerging and ceasing to exist. Nevertheless, that we perceive these representations as successive, and hence as part of the same temporal frame and as subject to certain laws, is only possible by means of a notion of time that ought to be independent from succession (and simultaneity).

References

ALLISON, H. "The Non-Spatiality of Things in Themselves for Kant". *Journal of the History of Philosophy*, v. 14, n. 3, p. 313-321. Baltimore: jul. 1976.

CAIRD, E. *The Critical Philosophy of Kant*. New York: Macmillan and Co, 1889.

CHABBOUH JUNIOR, M. A. *A aprioridade e a subjetividade de espaço e tempo na 'Estética Transcendental'*. São Paulo: Educ, 2014.

FALKENSTEIN, L. "Kant, Mendelssohn, Lambert and the Subjectivity of Time". *Journal of the History of Philosophy*, v. 29, n.2, p. 227-251. Baltimore: april 1991.

FUTCH, M. J. *Leibniz Metaphysics of Time and Space*. Boston: Springer, 2008.

HATFIELD, G. "Kant on the Perception of Space (and Time)". In: GUYER, P. (Ed.). *Cambridge Companion to Kant and Modern Philosophy*. New York: Cambridge University Press, 2006. p. 61-93.

KANT, I. "An Marcus Herz. 21. Febr. 1772". In: REICKE, R. (Ed.). *Kants Briefwechsel*. Berlin and Leipzig: Walter de Gruyter & Co, 1922.

_____. *Crítica da Razão Pura.* Trans. Manuela Pinto dos Santos e Alexandre Fradique Morujão. 5th Edition. Lisboa: Fundação Calouste Gulbenkian, 2001.

_____. *Critique of Pure Reason.* Trans. Norman Kemp Smith. 5. ed. Londres: Macmillan and Co, 1918.

_____. "Forma e princípios do mundo sensível e do mundo inteligível". Trans. Paulo L. Licht dos Santos. In: _____. *Escritos pré-críticos*. 1. ed. São Paulo: Editora Unesp, 2005.

_____. *Kritik der Reinen Vernunft*. Berlin e Leipzig: Walter de Gruyter & Co, 1911.

_____. "On the form and principles of the sensible and the intelligible world". In: GUYER, P.; WOOD, A. (Eds.). *Immanuel Kant Theoretical Philosophy 1755-1770*. Trans. David Walford. New York, 1992.

_____. "To Marcus Herz. February 21, 1772". In: ZWEIG, A. (Ed. and Trans.) *Immanuel Kant:* correspondence. New York: Cambridge University Press, 1999.

KEMP SMITH, N. *A commentary to Kant's critique of pure reason.* New York: MacMillan Company, 1923.

KITCHER, Patricia. *Kant's Transcendental Psychology*. New York: Oxford University Press, 1993.

LAMBERT J. H. "From Johann Heinrich Lambert. October 13, 1770". In: ZWEIG, A. (Ed. and Trans.). *Immanuel Kant:* correspondence. New York: Cambridge University Press, 1999.

_____. "Von Johann Heinrich Lambert. 18. Oct. 1770". In: REICKE, R. (Ed.). *Kants Briefwechsel*. Berlin e Leipzig: Walter de Gruyter & Co, 1922.

LEIBNIZ, G. W. *Correspondência com Clarke*. Trans. Carlos Lopes de Mattos. São Paulo: Abril Cultural, 1979.

MENDELSSOHN, M. "From Moses Mendelssohn. 25 December, 1770." In: ZWEIG, A. (Ed. and Trans.). *Immanuel Kant:* correspondence. New York: Cambridge University Press, 1999.

_____. "Von Moses Mendelssohn. 25. Dec. 1770". In: REICKE, R. (Ed.). *Kants Briefwechsel*. Berlin e Leipzig: Walter de Gruyter & Co, 1922.

PATON, H. J. *Kant's metaphysic of experience:* a commentary on the first half of the Kritik der reinen Vernunft. New York: MacMillan Company, 1936.

TRENDELENBURG, A. „Über eine Lücke in Kants Beweis von der ausschliessenden Subjektivität des Raumes und der Zeit". In: _____. *Historische Beiträge zur Philosophie*. Berlin: Verlag von G. Bethge, 1867, v. 3, p. 215-276.

Revista de Filosofia
ARGUMENTOS

Eduardo Simões*

Newton metafísico

RESUMO

O objetivo do presente artigo é o de apresentar o pensamento de Isaac Newton para além das fronteiras da Física experimental. Trata-se de um esforço em, mesmo conhecendo o gênio de objetividade do cientista inglês, demonstrar que a sua teoria escapa ao campo da Física para o da metafísica quando introduz em sua teoria gravitacional elementos indemonstráveis, tais como: espaço absoluto, tempo absoluto, éter, etc. O comportamento de Newton reflete uma prática comum às teorias científicas que, mesmo sem a admissão explícita da Ciência, ante da ausência da contraparte material da teoria, introduz elementos *ad hoc*, cuja função é a de manter a universalização e a identidade formal do sistema.

Palavras-chave: Newton; Metafísica; Espaço Absoluto; Tempo Absoluto; Éter.

ABSTRACT

The aim of this paper is to present the thought of Isaac Newton beyond the boundaries of experimental physics. It is an effort in, even knowing the genius of the English scientist objectivity, demonstrate that his theory escapes the field of physics to metaphysics when entering into its gravitational theory unprovable elements, such as absolute space, absolute time, ether, etc. The Newton's behavior reflects a common practice to scientific theories that, even without explicit admission of Science, compared to the absence of the material counterpart of the theory, introduces ad hoc elements, whose function is to maintain the formal and universal identity system.

Keywords: Newton; Metaphysics; Absolute Space; Absolute Time; Ether.

* Doutor em Filosofia pela Universidade Federal de São Carlos – UFSCar e professor da Faculdade de Ciência e Tecnologia de Montes Claros - Mg (FACIT). Email: eduardsimoes@yahoo.com.br.

Introdução

Conhecido como uma das mentes mais brilhantes da história da humanidade, Isaac Newton (1643 - 1727) foi imortalizado por sua obra mais significativa, o *Principia Mathematica* (*Princípios Matemáticos da Filosofia Natural*) de 1687. Nela ele consegue promover a unificação dos corpos planetários e terrestres por meio de um conjunto de equações capazes de prever exatamente – com base no volume de um corpo qualquer, na velocidade e na direção do movimento – como esse corpo se movimentaria sob o impulso de uma força conhecida. Com isso, postulou que se fosse dado a conhecer as posições e forças de todas as coisas no universo em um determinado instante, e se predissesse o curso integral dos acontecimentos desde os maiores corpos do universo aos mais leves átomos, nada seria incerto, e o futuro, à semelhança do passado, estaria presente diante de seus olhos.

Mas foi a descoberta da lei da gravidade o grande feito de Newton. Sua ideia foi a de que existe uma força invisível que exerce controle sobre a matéria sem haver um contato físico direto. A palavra gravidade foi cunhada a partir da palavra latina *gravitas*, que significa "peso". Com ela explicou com tanta precisão os movimentos das luas de Júpiter, de Saturno e da Terra, bem como os movimentos de todos os planetas ao redor do sol, que nos duzentos anos seguintes poucas melhorias significativas foram feitas em relação à sua obra. Essa força invisível está em ação entre as massas e é proporcional ao valor delas e inversamente proporcional ao quadrado da distância entre elas. Isso significa que, se duas massas são separadas, a força da gravidade entre elas diminui de tal forma que, quando a distância chega a 10 vezes, a força é de 100 vezes (quadrado de dez) menor do que a atração inicial. No caso do Sol, que está 400 vezes mais distante da Terra do que a Lua, a fator inversamente proporcional redutor da força gravitacional fica em cerca de 400^2 (16.000) – mas essa enorme redução é compensada pela massa imensamente maior do Sol em comparação à da Lua (a proporção de massa Sol-Lua é 30.000.000:1). Assim, a Terra continua orbitando o Sol. Toda essa explicação faz parte do terceiro livro[1] do *Principia* que termina por explicar os movimentos precisos da Lua e ensinar que as marés oceânicas se devem à atração gravitacional da Lua e do Sol sobre as águas. Além disso, calcula a atração do Sol sobre os cometas.

[1] O *Principia* que granjeou imediatamente uma fama para Newton, na verdade, é um livro muito complexo e difícil de compreender (cinquenta anos se passaram até que o esquema newtoniano fosse plenamente aceito e ensinado nas escolas e universidades). Ele se divide em três livros, embora tenha sido publicado em um único volume em 1687: o primeiro livro trata da mecânica e explica a razão porque os corpos se movem de determinada maneira no espaço vazio; o segundo livro trata do movimento dos corpos em meios que oferecem resistência, como o ar ou a água; e o terceiro livro é o que trata da estrutura e funcionamento do sistema solar e da gravidade.

Mas, é sabido que as correspondências entre Newton e Boyle eram frequentes e que seu pensamento, especialmente no que concerne a aceitação de um meio etéreo universal, teria sido influenciado por esse último[2]. Cabe-nos, portanto, introduzir o pensamento de Boyle no cenário da ciência moderna, levantar suas principais contribuições para o desenvolvimento da física clássica, para somente mais tarde, averiguar quem é o "Newton metafísico" – proposta principal de nosso artigo.

A metafísica de Boyle como resposta aos problemas da ciência moderna

Robert Boyle (1627-1691), foi um físico e químico irlandês que escreveu *O Químico Cético* (1661) onde defendeu o ideal de que as substâncias devem ser estudadas por meio de experiências práticas e que só são corretas as teorias comprovadas por experiências. Mesmo assim, foi ele responsável por uma importante construção metafísica explicativa da realidade na modernidade.

Sua teoria não traça os limites de sua atuação como químico ou como filósofo: aceita a visão mecânica cartesiana de mundo, valoriza as explicações qualitativas e teleológicas, insiste na realidade das qualidades secundárias (até então, combatidas por seus predecessores), mantém uma visão pessimista sobre o conhecimento humano e constrói sua estranha filosofia do éter. Tudo isso, tendo em vista que jamais perdera sua visão religiosa, onde Deus e o mundo mecânico mantêm uma relação de intimidade.

O ponto a ser destacado sobre trabalho de Boyle, e que serve diretamente aos nossos interesses, é que ele remonta ao atomismo que havia sido reintroduzido na ciência medieval e na modernidade, retomado por Gassendi e Descartes. No entanto, sua concepção de atomismo, ou de filosofia corpuscular, ou ainda, de filosofia mecânica, pretende ser apresentada subtraída as conotações metafísicas dos que o precederam.

> Supus poder prestar pelo menos um serviço não-desprezível aos filósofos corpusculares ilustrando algumas de suas noções com experimentos sensoriais e manifestando que as coisas por mim tratadas podem ser pelo menos plausivelmente explicadas sem recurso a formas inexplicáveis, qualidades reais, os quatro elementos peripatéticos, ou ainda os três princípios químicos. (BOYLE, 1672, VI. p. 356).

Sua proposta era a de analisar a química das coisas que nos rodeiam. Análise que fosse para além dos métodos místicos e mágicos da alquimia (que considerava o sal, o enxofre e o mercúrio como os três princípios químicos

[2] "Seu próprio pensamento sobre o assunto parece ter sido estimulado intimamente por Boyle, com quem tinha estreita comunicação a respeito de tais questões, como prova sua carta, datada de 1678, ao famoso químico." (BURTT, 1983, p. 149).

constituintes últimos da matéria), dos quatro elementos peripatéticos (água, fogo, terra e ar) e das concepções de átomo como entidade metafísica subjacente a toda realidade. Tratava-se de uma química fundamentada na análise racional dos fatos sensoriais e confirmada pela experiência.

Mesmo cheio de boas intenções, talvez pelas suas convicções religiosas, Boyle deixou-se atraiçoar quanto à sua fundamentação na experiência. Propôs a defesa de uma visão mecânica de mundo (mesmo que as fronteiras da mecânica até o presente momento ainda não estivessem totalmente delimitadas) onde a matemática (a metafísica matemática aos moldes de Galileu e Descartes) serviria à interpretação atomística do mundo (BURTT, 1983, p. 137). Sua concepção era a de que os princípios matemáticos eram "o alfabeto com que Deus escreveu o mundo":

> Encaro os princípios metafísicos e matemáticos [...] como verdades de tipo transcendental, que não pertencem propriamente seja à filosofia, seja à teologia, mas que constituem bases universais e instrumentos de todo o conhecimento que nós, mortais, podemos adquirir. (BOYLE, 1672, VI, p. 711).

A visão mecânica da natureza envolve, portanto, uma concepção mecânica de suas operações: quase todos os tipos de qualidades podem ser produzidos mecanicamente e, em última análise, os agentes corpóreos podem ser redutíveis a átomos, dotados apenas de qualidades primárias.

Das qualidades primárias, Boyle destaca especialmente o movimento e tenta explicar toda variedade e mudança através dele. É a partir da matéria, posta em movimento, que todos os fenômenos podem ser explicados (sejam eles os infinitamente grandes ou infinitamente pequenos).

> O movimento, que parece um princípio tão simples, especialmente nos corpos simples, pode, mesmo neles, ser muito diversificado; pois ele pode ser mais ou menos rápido em graus infinitamente variados; pode ser simples ou composto, uniforme ou variado, e a maior rapidez pode ocorrer no início ou no fim. O corpo pode mover-se em linha reta, ou circular, ou segundo alguma outra linha curva; [...] o corpo pode também ter movimento ondulante, [...] ou apresentar rotação ao redor de suas partes centrais, etc. (BOYLE, 1672, III, p. 299).

A explicação que Boyle dá do mundo a partir do movimento visa, na verdade, demonstrar que, pelas suas permutações e combinações, um número pequeno de diferenças primárias de movimento, figura, volume pode dar origem, a partir de várias combinações possíveis, a uma grande diversidade de fenômenos. E esses movimentos, assim como pensavam Galileu e Descartes, deveriam ser explicados em termos matemáticos exatos.

Mas, se até agora Boyle parece apresentar uma concepção coerente de realidade, explicada em termos matemáticos, qual é a sua contribuição para

uma metafísica explicativa da realidade? É justamente nesse ponto que pretendíamos chegar. Ao propor a explicação dos fenômenos a partir das qualidades primárias do movimento, do volume e da forma, Boyle não conseguiu fugir das famigeradas qualidades secundárias e, para elas, não consegue explicações que se esquivassem das dos peripatéticos:

> não se deve desprezar as explicações em que efeitos particulares são deduzidos a partir das mais óbvias e familiares qualidades ou estados dos corpos, tais como o calor, o frio, o peso, a fluidez, a dureza, a fermentação, etc. (BOYLE, 1672, I, p. 308).

E o que ele faz, além de confirmar a realidade das qualidades secundárias, é reafirmar uma posição na qual se mantém fiel a antiga noção de causa final – todas as qualidades apontam para algo que transcendentalmente as antecede: existe "a admirável cooperação das diversas partes do universo para a produção de efeitos particulares; e é difícil dar explicações satisfatórias para todos eles sem reconhecer um ser inteligente que crie ou disponha das coisas." (BOYLE, 1672, II, p. 76).

A adesão de Boyle a conceitos que pareciam ter sido superados pelos seus predecessores (Galileu, Descartes, etc.) deve-se única e exclusivamente à sua necessidade de resgatar o homem do materialismo do século XVII. O mundo real era o domínio dos pensamentos de Galileu e Descartes; esse mundo era matemática e mecanicamente inteligível e todo esforço racional devia-se a explicação do seu funcionamento. A razão tornou-se, então, o fundamento último de sua explicação. No entanto, essa visão que dominou a época, esqueceu-se do homem e o colocou como uma espécie de apêndice, puro espectador da natureza.

> Contrapondo-se a essa tendência aparentemente irresistível de expulsar o homem da natureza e de diminuir sua importância, Boyle empenhou-se positivamente em reafirmar o lugar factual do homem no cosmos e sua dignidade singular como filho de Deus. (BURTT, 1983, p. 142).

E é por isso que as qualidades primárias não são mais *reais* que as secundárias: elas estão no homem e, "uma vez que o homem, com seus sentidos, é parte do universo, *todas* as qualidades são igualmente reais". E, como ele próprio afirma,

> não vejo a necessidade de que a inteligibilidade com relação ao entendimento humano seja necessária para a verdade ou a existência de uma coisa, assim como a visibilidade com relação ao olho humano não é necessária para a existência de um átomo, ou de um corpúsculo de ar, ou dos eflúvios de um imã, etc. (BOYLE, 1672, IV, p. 450).

E, justamente pensando nessa noção de não-necessidade da inteligibilidade das coisas para que elas de fato existam, é que Boyle propõe uma das

mais estranhas (no entanto, muito comum em sua época) concepções metafísicas da história da química moderna: a filosofia do éter.

Como se disse, na época de Boyle era muito comum a crença na existência de um meio etéreo – seja para justificar a comunicação do movimento por impacto sucessivo ou através das distâncias (Descartes), ou para explicar os fenômenos do magnetismo. Ele próprio, num primeiro momento, encarou-o como algo duvidoso, mas *a posteriori* admitiu que pudesse sim existir uma sustância etérea "muito tênue e difusa".

> Considerei que a parte interestelar do universo, consistente de ar e de éter, ou de fluidos análogos a um deles, é diáfana; e que o éter é como se fosse um vasto oceano no qual os globos luminosos, que, aqui e ali, nadam como peixes, por seus próprios movimentos, ou que, como corpos em redemoinhos, são transportados pelo ambiente, encontram-se grandemente dispersos, de modo que a proporção das estrelas fixas e dos corpos planetários com relação à parte diáfana é extremamente pequena e mal pode ser considerada. (BOYLE, 1672, IV, p. 451).

Essa "substância", em Descartes, era concebida como um fluido homogêneo e fleumático que preenchia todo o espaço (com uma série de vórtices de diversos tamanhos) não ocupado por outros corpos e que não possuía características que não pudessem ser deduzidas da extensão. Em Boyle o éter mostraria sua serventia na medida em que pudesse encontrar nele dois tipos de matéria: uma que explicasse a comunicação por movimento e a outra que justificasse os fenômenos do magnetismo. E ele vai encontrar justamente na teoria corpuscular a orientação de que precisava para comungar essas duas dificuldades da ciência moderna em uma espécie de filosofia do éter. Diz:

> Pode, portanto, não ser desarrazoado confessar-vos que entretive leves suspeitas de que, além dos tipos mais numerosos e uniformes de *partículas diminutas* de que alguns dos novos filósofos pensam que é composto o éter sobre o qual venho discorrendo, é possível a existência de outros tipos de *corpúsculos*, capazes de consideráveis operações quando encontram corpos congruentes sobre os quais podem atuar; mas, embora seja possível, e talvez provável, que os efeitos que estamos considerando possam ser explicados plausivelmente pelo éter, tal como ele é realmente entendido, tenho certas suspeitas de que tais efeitos possam não ser devidos exclusivamente às causas que lhes são imputadas, mas sim que possivelmente existam, como eu começava a dizer, tipos peculiares de *corpúsculos*, que até aqui não tem nome próprio, que podem revelar faculdade e maneiras de atuar peculiares ao encontrar-se com corpos cuja estrutura os leve a admitir a eficácia desses agentes desconhecidos ou a concorrer para ela. Esta minha suspeita parecerá menos improvável se considerardes que, embora no éter dos antigos não existisse nada que se pudesse notar além de uma substância difusa e muita tênue, hoje estamos dispostos a admitir que existe permanentemente no ar uma multidão de eflúvios que se movem em um curso determinado entre o pólo norte e o pólo sul. (BOYLE, 1672, III, p. 316, grifos nossos).

A introdução do éter na teoria boyliana segue o mesmo princípio de seus antecessores, salvo acréscimos sutis no que concerne ao encontro com os corpos congruentes sobre os quais podem atuar, isto é, trata-se de uma teoria corpuscular, onde o éter é composto dos tipos mais numerosos e uniformes de partículas diminutas.

Certo é que, mesmo sem um aparato na experiência, a concepção de éter serviu para preencher duas funções distintas sobre as quais os modernos debatiam: a explicação da propagação do movimento através de distâncias e a explicação de fenômenos tais como a coesão, o magnetismo, etc. que, até então, não podiam ser reduzidos à matemática exata. E essa "distinção entre dois tipos de matéria etérea, feita com o objetivo de que o éter pudesse fornecer uma explicação adequada para estes dois tipos de fenômenos, será novamente examinada com Newton." (BURTT, 1983, p. 149).

Quando fala em éter como uma necessidade de um princípio explicativo de algo que até então era inexplicável, Boyle se debate com o mesmo problema de seus antecessores, que é o de responder: o que é o éter? De que é composto? Com qual matéria do universo ele se identifica? E para tais perguntas, dado a impossibilidade concreta de resposta, ele se vê, também, obrigado a recorrer à "realidade" do átomo como válvula de escape para resolução de seus problemas teóricos, ainda que não houvesse qualquer prova empírica irrefutável da existência física dos átomos. E, mais uma vez, a metafísica se sobrepõe a uma explicação que se pretendia química dos fenômenos. É Newton quem assumirá o compromisso da explicação da realidade onde todas as hipóteses seriam eliminadas, restando somente a experiência para confirmar os dados da natureza.

Os componentes metafísicos da física newtoniana

Uma das principais preocupações do pensamento de Newton foi com busca de uma resposta à pergunta sobre como se altera o estado de movimento de uma massa puntiforme (que tem forma ou aparência de ponto) num tempo infinitamente curto sob a influência de uma força externa. Para essa questão, ele chegou à resposta analisando a trajetória de uma partícula ideal. Aplicou as suas leis do movimento a um pequeno intervalo de tempo e, com isso, previu a posição da partícula e a velocidade ao final desse intervalo. E essa experiência, que foi repetida sucessivas vezes aplicando o mesmo cálculo, permitiu-lhe estimar a trajetória total. E isso só foi possível com a aplicação de um atalho matemático que ele inventou (paralelamente a Gottfried Leibniz) chamado cálculo diferencial. Com o cálculo ele conseguiu abreviar o processo passo a passo o que lhe possibilitou analisar o que acontece à velocidade de uma partícula em movimento à medida que a diferença temporal se torna infinitesimal. Nisso resultou as suas três conhecidas leis do movimento:

a) a primeira diz que "todo corpo persevera em seu estado de repouso ou de movimento retilíneo uniforme, a menos que seja compelido a mudar seu estado por forças aplicadas". Em outras palavras, *um corpo continuará em repouso a menos que uma força atue sobre ele, e um corpo em movimento retilíneo uniforme continuará a mover-se na mesma velocidade em linha reta a menos que uma força atue sobre ele.* Isso quer dizer que, uma bola em uma superfície plana perfeita somente se moverá se uma força atuar sobre ela. Se uma força a faz começar a rolar e se ela não encontra nenhum atrito com a superfície ou algum obstáculo em seu caminho, ela continuará rolando na mesma direção para sempre. Esse princípio pode também ser chamado princípio da inércia, sendo esta a propriedade da matéria que a faz resistir a qualquer mudança em seu movimento;
b) a segunda lei diz que "uma alteração no movimento é proporcional à força motora e ocorre ao longo da linha reta na qual tal força é aplicada". O que quer dizer que *a aceleração (taxa de variação do movimento[3]) é diretamente proporcional à força*. Por exemplo, quanto maior a força gerada pelo motor de um automóvel, mais o carro se acelerará. O dobro da força duplicará a aceleração;
c) no caso da terceira lei, essa diz que "para qualquer ação existe sempre um reação oposta e idêntica; em outras palavras, as ações de dois corpos um sobre o outro são sempre idênticas e sempre opostas em termos de direção". Por exemplo, a "ação" de uma bala disparada por um revolver, resulta na "reação" do coice da arma. Ou então, que quando estamos sentados em uma cadeira, esta exerce uma força para cima de nós para compensar o nosso peso, que pressiona para baixo. Dizia Newton que isso acontece também no céu: enquanto a Terra exerce um arranjo gravitacional sobre a Lua, mantendo-a em órbita, a Lua faz o mesmo em relação à terra, criando as marés nos oceanos.

O ponto fraco da teoria gravitacional de Newton concentra-se, no entanto, na exigência promovida pela mesma, da existência de um tempo e espaço absolutos. E é justamente esse o rito de passagem de sua física para a metafísica.

É sabido de todos a obsessão de Newton pela conclusão experimental de suas teorias. Tanto é que somente vinte anos depois de ter chegado a todas as conclusões do *Principia*, encorajado pelo matemático Edmond Halley (1656-1742) que arcou com os custos da publicação, tais conclusões chegaram a público. Sua justificativa era a de que para as suas descobertas seriam necessárias mais experimentações e provas. Determinados cálculos não lhe pareciam precisos, pois eram baseados no valor aceito (mais incorreto) do diâmetro da Terra e ele não admitia hipóteses. "Se ainda houver alguma dúvida [sobre minhas conclusões], é melhor colocar o caso em circunstâncias mais aprofundadas do experimento do que aquiescer à possibilidade de qualquer explicação hipotética." (NEWTON, *Opera*, 1779 *apud* BURTT, 1983, p. 173). Isso porque,

[3] A "quantidade de movimento", ou "movimento", é dada pelo produto da massa de um corpo por sua velocidade: "F=ma".

> Qualquer coisa não deduzida de fenômenos deve ser chamada de hipótese; e hipóteses, sejam metafísicas ou físicas, referentes a qualidades ocultas ou mecânicas, não têm lugar na filosofia experimental. Nesta filosofia, proposições particulares são inferidas dos fenômenos, e tornadas gerais, em seguida, por indução. Assim foi que a impenetrabilidade, a mobilidade, e a força impulsiva dos corpos, e as leis de movimento e de gravitação foram descobertas. (NEWTON, *Principles*, III, 1803, p. 314).

Mesmo com tantas reservas com relação às hipóteses, suas concepções sobre espaço e tempo, especialmente sobre espaço e tempo absolutos, deixam margens para questionamentos, principalmente sobre o valor não-hipotético dos mesmos. E é justamente nesse ponto que sua teoria se inicia nas concepções metafísicas da ciência moderna.

Apesar dos avanços de seus predecessores, foi com Newton que a natureza passou a ser pensada essencialmente como o domínio das massas que se movem de acordo com leis matemáticas no espaço e no tempo sob a influência de forças definidas e confiáveis. A definição de massa é dada por ele já no primeiro parágrafo do *Principia* e é feita em termos de densidade e volume. E a descoberta sobre ela é que a mesma tem diferentes pesos a distâncias diferentes do centro da Terra e que é composta em última análise de *partículas* absolutamente rígidas, indestrutíveis, impenetráveis, etc. E todas as mudanças na natureza devem ser vistas como separações, associações e movimentos desses *átomos* permanentes que são predominantemente matemáticos.

> Aprendemos, pela experiência, que a maior parte dos corpos é dura; e como a dureza do todo deriva da dureza das partes, nós justamente inferimos, portanto, a dureza das partículas não divididas não somente dos corpos que percebemos, mas também de todos os outros. Não é da razão, mas, sim, da sensação que concluímos que todos os corpos são impenetráveis [...]. E daí concluímos serem as menores partículas de todos os corpos também dotadas de extensão, duras, impenetráveis, capaz de serem movimentadas e dotadas de suas próprias *vires inertiae.* (NEWTON, *Principles,* 1803, II, p. 161).

Vemos aqui que Newton também recorre à realidade do átomo, até então desconhecido empiricamente, para explicar a composição última da matéria. Nesse momento, ainda é admissível, mesmo que por dedução, o emprego do atomismo: é "óbvio" que por trás de todo real deve haver um componente último do mesmo real; que aquilo que caracteriza o todo deve caracterizar também a parte. Portanto, aqui, ainda é possível conceber os argumentos newtonianos como genuinamente físicos e manter o devido respeito a sua personalidade experimental. Mas, as coisas se complicam na medida em que ele passa da definição de massa à definição de tempo e espaço absolutos – é nesse ponto que ele abandona seu empirismo não conseguindo se esquivar da metafísica. Ele mesmo admite que ao oferecer caracterizações de espaço, tempo, e movimento, "devemos abstrair-nos dos nossos sentidos e considerar as coisas por si próprias,

distintas do que são apenas medidas perceptíveis delas." (NEWTON, *Principles,* 1803, I, p. 9). As definições abaixo são retiradas de Burtt (1983, p. 193-194):

> I – O tempo absoluto, verdadeiro e matemático, por si, e pela sua própria natureza, flui uniformemente, sem observar qualquer coisa externa, e é chamado, também, de duração: o tempo relativo, aparente e comum, é uma medida perceptível e externa (seja precisa ou variável) de duração por meio do movimento, que é comumente utilizada em vez do tempo verdadeiro, como uma hora, um dia, um mês, um ano.
> II – O espaço absoluto, por sua própria natureza, indiferente a qualquer coisa externa, permanece sempre similar e imóvel. O espaço relativo é uma dimensão móvel ou medida dos espaços absolutos; o que nossos sentidos determinam por sua posição relativa aos corpos, e que é vulgarmente tido como espaço imóvel; esta é a dimensão de um espaço subterrâneo, aéreo ou celeste, determinada por sua posição com relação à Terra. O espaço absoluto e o relativo são iguais em figura e magnitude; mas não permanecem sempre numericamente iguais. Porque, se a Terra se move, por exemplo, um espaço do nosso ar que, com relação à Terra, sempre permanece o mesmo, será em determinado momento parte do espaço absoluto no qual passa o ar; em outro momento, corresponderá a outra parte do mesmo, e assim, absolutamente compreendido, será perpetuamente mutável.

A confusão acaba de ser instaurada: o que é o tempo absoluto? E o relativo? E quanto ao espaço absoluto? E o relativo? Qual é a necessidade subjacente a essas divisões? O que as justifica? Todas essas respostas são dadas pelo próprio Newton, encaixam perfeitamente bem em seu sistema, mas, parece-nos a contragosto da própria realidade empírica.

"O tempo absoluto, verdadeiro e matemático, por si, e pela sua própria natureza, flui uniformemente [...]". "O espaço absoluto, por sua própria natureza, indiferente a qualquer coisa externa, permanece sempre similar e imóvel [...]". Vejamos um exemplo de sua justificativa: um passageiro de um barco se move em relação ao barco, o barco se move em relação à Terra, a Terra se move em relação ao Sol – e tudo o que é físico se move em relação a um referencial espaço-temporal que se encontra em "repouso", absoluto. Quanto ao espaço e tempo absolutos

> estes são infinitos, homogêneos, entidade contínuas, inteiramente independentes de qualquer objeto perceptível ou movimento pelo qual tentamos medi-lo, e o tempo flui uniformemente da eternidade para a eternidade, e o espaço todo, ao mesmo tempo, em imobilidade infinita (BURTT, 1983, p. 195).

A questão que ora nos fica é: qual é a natureza deste referencial universal?

Ainda sem respostas para as confusas elucubrações de Newton, vem-nos imediatamente à mente a questão de saber se a exigência de tempo e espaço absolutos convive com a concepção de um movimento absoluto ou mesmo um repouso absoluto? E o que seriam eles? A resposta é positiva.

Quando um corpo se transfere de uma parte do espaço absoluto para outra parte, temos um *movimento absoluto* e quando há uma continuidade de um corpo na mesma parte do espaço absoluto temos o *repouso absoluto*.

> A existência de um movimento absoluto implica a existência de um *ambiente infinito* no qual podem mover-se, e a mensurabilidade exata daquele movimento sugere que *esse ambiente é um sistema geométrico perfeito e um tempo matemático puro – em outras palavras, movimento absoluto sugere duração absoluta e espaço absoluto.* (BURTT, 1983, p. 202).

O que vemos aqui é que Newton, forçosamente, quer transformar tempo e espaço em entidades reais e absolutas que existem independentemente da mente humana, onde o movimento funciona na mais perfeita harmonia. Essa certeza proporcionou uma fundação sólida sobre a qual a ciência construiu o que veio chamar de "física clássica", que durou dois séculos, e que funcionou perfeitamente bem até o advento da relatividade no século XX. Só que sua teoria se aplica bem ao movimento dos grandes sistemas; permite que uma inteligência humana, se lhe fosse dado conhecer as posições e forças das coisas no universo em um determinado instante, prediga o curso integral dos acontecimentos, desde os maiores corpos do universo aos mais leves átomos – desde que seus movimentos sejam harmônicos.

Alguns religiosos de plantão, como era o caso Leibniz, por exemplo, que foi um crítico ferrenho de Newton, apontaram para aquilo que chamaram de influência anticristã dos *Principia*: as posições fundamentais foram as de que espaço e tempo infinitos e absolutos, eram admitidos como entidades independentes, vastas, nas quais as *massas moviam-se mecanicamente*, e isso significaria dar a Deus férias de suas funções primordiais. Onde caberia a ação divina se tudo funcionasse como uma espécie de relógio, harmonicamente acertado? Deus parecia ter sido varrido da existência e nada havia para tomar o seu lugar exceto esses seres matemáticos ilimitados. Isso ecoou mais intolerável para Newton do que a própria querela entre ele e Leibniz sobre o plágio que esse último teria feito de sua invenção: o cálculo diferencial[4]. Mas, as acusações eram injustificadas. Esse relógio que era o universo, para Newton, não poderia funcionar para sempre sem a intervenção de Deus, pois, sendo assim, sua necessidade seria supérflua. Certas irregularidades no sistema solar, não explicadas pelos movimentos dos planetas, poderiam sim tirar todo o sistema dos eixos, daí caberia a intervenção divina para colocar tudo novamente em ordem.

[4] Sobre a intriga entre Newton e Leibniz sobre quem teria antecedido na invenção do cálculo uma boa referência é a seguinte: HELLMAN, Hal. *Grandes debates da ciência:* dez das maiores contendas de todos os tempos. Tradução José Oscar de Almeida Marques. São Paulo: Editora Unesp, 1999.

Sua outra concepção sobre Deus é a de que Ele é o *sensorium uniforme e ilimitado*, onde todos os corpos se movem. Ele é o próprio espaço absoluto.

> É admitido por todos que o Supremo Deus existe, necessariamente; e pela mesma necessidade ele existe *sempre* e *em toda parte*. Donde ele também é todo similar, todo olho, todo ouvido, todo cérebro, todo braço, todo poder de percepção, para compreender e para agir; mas de maneira não-humana, não-corpórea; de maneira absolutamente desconhecida por nós. (NEWTON, *Principles,* 1803, II, p. 311).

Essas duas concepções acima (de Deus como coordenador do funcionamento da máquina e de Deus como *sensorium*), mais uma vez, foi motivo de escárnio por parte de Leibniz: primeiramente, "ria-se da suposição de que Deus seria uma espécie de encarregado de manutenção em nível astronômico", segundo, quanto à ideia de que o espaço era uma espécie de *sensorium* de Deus, o questionamento de Leibniz era: "Será que Deus precisaria de órgãos sensoriais a fim de perceber?." (HELLMAN, 1999, p. 85-86). Certo é que Deus permanece intacto em seu sistema e que as concepções newtonianas muito além de físicas, estão carregadas de uma robusta metafísica que as sustentam e as mantém.

James Gleick, um biógrafo de Newton, diz que "Deus inspirou a crença de Newton em um espaço absoluto e um tempo absoluto", mesmo assim, ele deve ter tido algumas dúvidas sobre a veracidade de um tempo e espaço absolutos, pois também observou em *Principia*:

> talvez não exista um movimento uniforme que possa servir para mensurar com precisão o tempo. Talvez nenhum corpo esteja efetivamente em repouso de modo a servir de referência para a posição e o movimento de outros (NEWTON, *Principles*, 1803, II, p. 315).

Para o jovem estudante de física Einstein, uma especulação similar funcionaria como forte estímulo para a criação da teoria da relatividade.

Mas, a presença de premissas teológicas na física newtoniana sobre espaço e tempo, é mais reforçada na medida em que aparece um aspecto fortemente conservador em sua metafísica: Newton concebe que exista um meio etéreo suscetível a vibrações.

Na época de Boyle, o meio etéreo era utilizado para justificar o movimento propagado à distância e explicava fenômenos extra mecânicos como eletricidade, magnetismo e coesão. Em Descartes aparece como fluido denso, compacto, que equilibrava os planetas em suas órbitas pelo seu movimento de vórtices. Já em Newton, cujo pensamento a esse respeito havia sido estimulado por Boyle, sua concepção sobre o meio etéreo que a princípio soava como uma hipótese, passou a ser um elemento fundamental de sua metafísica (lembre-se de como ele atacava qualquer hipótese):

> Se tivesse de presumir uma hipótese, seria esta, se proposta de forma mais geral, de modo a não determinar o que é a luz; além de ser ela algo

> capaz de estimular vibrações no éter; pois assim ela tornar-se-á geral e abrangerá outras hipóteses, de modo a deixar pouco espaço para invenção de novas hipóteses (BREWSTER, 1851, p. 390 *apud* BURT, 1983, p.214).[5]

E, com essa hipótese, Newton passa a explicar vários tipos de fenômenos como a gravidade, a eletricidade, a coesão, a sensação animal e o movimento, a refração, a reflexão e as cores da luz, etc.

> Assim, a atração gravitacional da Terra pode ser causada pela contínua condensação de um outro espírito etéreo similar, que não o corpo fleumático principal do éter, mas algo muito tênue e difundido sutilmente através dele, de natureza talvez oleosa, pegajosa, tenaz e elástica, e desempenhando uma relação com o éter muito semelhante à que o espírito aéreo vital requer para a conservação da chama e que os movimentos vitais fazem ao ar. (BREWSTER, 185I, p. 393-394 *apud* BURTT, 1983, p. 213-214).

E no último parágrafo de *Principia*, onde Newton já havia superado a divisão entre o corpo fleumático principal do éter e os diversos espíritos etéreos difundidos através dele, escreve:

> Agora acrescentaremos algo concernente a um certo espírito muito tênue, que permeia e permanece escondido em todos os corpos densos, por cuja força e ação as partículas dos corpos atraem-se mutuamente a distância próximas e se integram, se contíguas; e os corpos elétricos operam, a maiores distâncias, tanto repelindo como atraindo os corpúsculos vizinhos; e a luz é emitida, refletida, refratada, desviada e aquece os corpos; e toda sensação é estimulada, e os membros dos corpos animais se movem ao comando da vontade, pelas vibrações desse espírito, propagado mutuamente ao longo dos filamentos sólidos dos nervos, dos órgãos externos de sensação ao cérebro, e do cérebro aos músculos. Mas essas são coisas que não podem ser explicadas em poucas palavras nem estamos providos de experimentos suficientes, necessários para uma determinação e uma demonstração acuradas das leis pelas quais esse espírito elétrico e elástico opera (NEWTON, *Principles*, 1803, II, p. 314).

Como se vê, Newton não possuía quaisquer certezas acerca dessa entidade "fantasmagórica" e, daí, podemos levantar alguns problemas a partir de suas palavras: o primeiro diz respeito às suposições que envolvem a explicação da gravidade. A todo o momento encontramos expressões como "assim *talvez* o Sol..." ou "quem quiser também pode *supor*...", e outras mais; e isso implica a falta de respostas conclusivas do próprio Newton para esse fenômeno. Sendo assim, fica mais fácil e universalizante deduzir a presença de um "espírito etéreo" em "um corpo fleumático" – isso propicia uma independência formal ao seu sistema. O segundo problema fica por parte das dificuldades conceituais que geram sua teoria: o que seria esse corpo fleumático pelo qual

[5] Carta a Oldenburg, secretário da Sociedade Real, em 1675. Esta carta encontra-se no reunido de cartas de Brewster (em I, p. 390), *Memoirs of the Life, Writings and Discoveres of Isaac Newton*, Edinburgo, 1855 – citado por BURTT, 1983, p. 211.

espírito etéreo se move? E o que é o próprio espírito etéreo? Como foi visto na citação supramencionada, Newton não "provia de experimentos suficientes" que apresentassem respostas conclusivas para essas questões. Ele é mais um representante da herança daqueles que, não tendo aparatos técnicos e tecnológicos para explicar experimentalmente tais fenômenos, foi obrigado a recorrer à metafísica na explicação da física.

Só nos resta, por fim, tentar explicar a composição desse meio etéreo tal como Newton o concebe e essa explicação não poderia ser outra, para a nossa "surpresa", que não atomista:

> E se todos supusessem que o éter (como o nosso ar) pode conter partículas que tendem a afastar umas das outras (pois não sei o que é esse éter), e que suas partículas são extremamente menores que as do ar, ou mesmo que as da luz: a extrema pequenez de suas partículas pode contribuir para a grandeza da força pela qual aquelas partículas podem afastar-se umas das outras, e, desse modo, tornar aquele meio extremamente mais rarefeito e elástico que o ar, e, por consequência, extremamente menos capaz de resistir aos movimentos de projéteis e extremamente mais capaz de fazer pressão sobre os corpos volumosos, na sua tendência à expansão. (NEWTON, *Opticks*, 1721, p. 323).

Vemos que o éter de Newton tem a mesma natureza do ar, mas é muito mais rarefeito. Suas partículas são muito pequenas e estão presentes em maior quantidade de acordo com sua distância dos poros interiores dos corpos sólidos. São elásticas por possuírem poderes mutuamente repulsivos e tendem constantemente a afastar-se umas das outras e essa tendência é a causa dos fenômenos de gravidade. Todo o mundo físico pode consistir de partículas que se atraem em proporção ao seu tamanho, passando a atração através de um ponto zero para a repulsão até chegar às menores partículas que compõem o que se denomina éter. E essas são as consequências da metafísica newtoniana.

Conclusão

O impacto das teorias newtonianas ainda se faz sentir no século XX em muitos campos da ciência. A teoria das ondas luminosas usa as leis do movimento de Newton, e o mesmo se pode dizer da teoria cinética do calor. A teoria newtoniana foi importante também no desenvolvimento de nossa compreensão da eletricidade e do magnetismo e nas descobertas de Faraday e Maxwell em eletrodinâmica e óptica. Sua física norteou a ciência por mais de duzentos anos – até a primeira metade do século XX, quando Einstein demonstrou que a Física precisava crescer para além da estrutura newtoniana. E o necessário crescimento da Física a partir de Newton devia-se, também, pela necessidade de superação de elementos metafísicos incorporados no seu pensamento físico clássico. Se isso será realizado a história testemunhará.

Referências bibliográficas

BOYLE, Robert. *The works of the honourable Robert Boyle*. Londres: Ed. Thomas Birch, 1672. 6 v.

BURTT, Edwin A. *As bases metafísicas da ciência moderna*. Tradução de José Viegas Filho, Orlando Araújo Henriques. Brasília: Editora Universidade de Brasília, 1983.

HELLMAN, Hal. *Grandes debates da ciência:* dez das maiores contendas de todos os tempos. Tradução de José Oscar de Almeida Marques. São Paulo: Editora Unesp, 1999.

NEWTON, Isaac. *Isaaci Newtoni opera quae exstant omnia.* Edição Samuel Horsley, 5 vls., L.L.D: Londres, 1779.

_____. *Mathematical principles of natura philosophy.* Tradução de Andrew Motte. 3 vls. Londres, 1803.

_____. *Opticks*: or, a treatise of the reflections, refractions, inflection, and colours of light. 3 ed. Londres, 1721.

Revista de Filosofia
ARGUMENTOS

Tiaraju Andreazza*

Teoria moral e equilíbrio reflexivo

RESUMO

John Rawls afirma, em *A Theory of Justice*, que a função do teórico moral é formular princípios de justiça que caracterizam a nossa competência de distinguir entre o certo e o errado. O objetivo deste artigo é discutir o significado dessa afirmação. Analiso e argumento contrariamente à interpretação recentemente defendida por Mikhail, de acordo com a qual Rawls estaria propondo uma "concepção naturalística" de teoria moral como uma investigação empírica. Defendo que essa interpretação de Mikhail está baseada em uma compreensão equivocada da ideia de equilíbrio reflexivo.

Palavras-chave: Juízos morais ponderados; Teoria moral; Equilíbrio reflexivo.

ABSTRACT

John Rawls claims, in *A Theory of Justice,* that the moral theorist's role is to formulate principles of justice that characterize our competence to distingish between right and wrong. The aim of this article is to discuss the meaning of this claim. I analyze and argue against the reading recently advocated by Mikhail. According to Mikhail, Rawls is proposing a "naturalistic conception" of moral theory as a empirical inquiry. I maintain that this reading is based on a misreading of the idea of reflective equilibrium.

Keywords: Considered moral judgments; Moral theory; Reflective equilibrium.

* Doutorando em Filosofia do Programa de Pós-Graduação em Filosofia da Universidade do Vale do Rio dos Sinos (UNISINOS). Bolsista PROSUP/CAPES. Email: tiaraju.andreazza@gmail.com

Considerações iniciais

Na seção 9 de *A Theory of Justice* (*TJ*, daqui para frente), Rawls elabora uma polêmica analogia com uma teoria linguística a fim de explicar a natureza da teoria moral e o papel do teórico moral. Por meio dessa analogia ele esclarece que a função do teórico moral éformular princípios de justiça que caracterizam a nossa competência de distinguir entre o certo e o errado, ou princípios que "descrevem o nosso senso de justiça", assim como a função de um linguista écaracterizar a nossa capacidade de reconhecer sentenças bem-formadas formulando um conjunto de regras gramaticais que façam as mesmas discriminações que nós fazemos (RAWLS, 1999, p. 41). Recentemente Mikhail defendeu a interpretação de que em *TJ,* e com essa analogia,Rawls estaria simplesmente estendendo a teoria linguística de Chomsky para o campo da filosofia moral, com isso esboçando as bases de um programa de acordo com o qual os teóricos morais deveriam passar a adotar uma "concepção naturalística"tanto do seu objeto de estudo (no caso, a competência moral ou o senso de justiça) quanto da sua metodologia (MIKHAIL, 2011). O objetivo deste artigo éo de discutir a intepretação de Mikhail e de analisar as suas implicações. Como Rawls esclarece o que significa falar em "descrição"de um senso de justiça através da ideia de equilíbrio reflexivo, e como esta noção estáno centro da interpretação oferecida por Mikhail, este étambém um artigo que discute a natureza dessa ideia.

O meu objetivo é defender que a interpretação de Mikhail não é uma interpretação fiel de *TJ* porqueela depende de uma interpretação equivocada da ideia de equilíbrio reflexivo. Defenderei também que a interpretação de Mikhail resulta em uma compreensão equivocada do problema normativo de justificar princípios de justiça, uma compreensão que atribui uma autoridade ao teórico moral que Rawls não subscreve. O artigo é dividido em três seções. Na primeira seção analiso a ideia de equilíbrio reflexivo e como Mikhail a caracteriza. Na seção seguinte argumento por que essa interpretação não se ajusta ao texto de *TJ*, e, por fim, concluo na terceira seção defendendo por que ela acaba resultando em uma compreensão equivocada do problema justificacional.

Equilíbrio reflexivo e a interpretação de Mikhail

Na seção de *A Theory of Justice* (*TJ*, daqui por diante) dedicada à apresentar a sua concepção de teoria moral, Rawls indica em uma nota de rodapé que está seguindo o "ponto de vista geral" do artigo "Outline of a Procedure for Ethics" (*Outline,* daqui em diante), um artigo de 1951 em que ele sumariza a sua tese de doutoramento. Mikhail conta com o *Outline* como parte da sua evidência de um programa "linguístico" em *TJ*. Nós devemos começar então com uma breve análise do ponto de vista geral desse artigo.

O *Outline* é um ambicioso artigo no qual Rawls esboça uma solução para dois problemas diferentes. O primeiro problema é o de formular princípios justificáveis que, em casos em que há conflito de interesses, podem ser usados para determinar a quais interesses é correto ou justo dar preferência (RAWLS, 1951, p.178). O segundo objetivo, que podemos convencionar chamar de "epistemológico", é o de descrever um procedimento de escolha através do qual é possível mostrar que esses princípios são justificados (RAWLS, 1951, p. 183). Esse procedimento, que no que se segue eu gostaria de descrever brevemente, passou a ser conhecido na literatura pelo nome de equilíbrio reflexivo estreito (*narrow reflective equilibrium*).

O procedimento funciona do seguinte modo. Rawls parte da assunção de que as pessoas têm a capacidade para saber o que é certo e errado do mesmo modo como elas têm para saber o que é verdadeiro e falso, mas em geral nas situações concretas cotidianas essa capacidade não produz juízos morais confiáveis devido a estar sujeita a uma série de fatores distorcivos. Rawls estipula então uma lista de condições que um juízo moral deve satisfazer para que ele possa ser considerado confiável, defendendo, em resumo, que apenas os juízos morais ponderados (juízos realizados sob circunstâncias favoráveis ao exercício do juízo, quando não se está sob forte estresse emocional etc) de juízes morais competentes (agentes morais com um nível básico de inteligência, bem-informados dos fatos relevantes, dispostos a considerar o mérito e interesses de todos os envolvidos etc) são confiáveis[1]. Ele sustenta então que para demonstrar que os princípios são justificados nós precisamos mostrar que de alguma forma eles "explicam" essa classe de juízos, em que um princípio explica esses juízos se a conscienciosa aplicação do princípio para solucionar um caso particular levaria o agente moral, pelo uso desse princípio, a formar os mesmos juízos morais ponderados que os juízes morais competentes (RAWLS, 1951, p. 184-185). Rawls classifica esse procedimento como uma "investigação empírica." (RAWLS, 1951, p. 184).

Assim, um princípio justificado é um princípio que figuraria nessa explicação. Se nós quisermos descobrir se um juízo moral particular é justificado, nós devemos nos perguntar se ele poderia ser explicado por um conjunto de princípios que faria parte de tal explicação. Conforme Mikhail corretamente salienta, o *Outline* defende uma forte confluência entre as esferas descritiva (que busca por explicação) e normativa (que busca por justificação): a solução para um problema descritivo (explicar a capacidade moral de distinguir entre o certo e errado) acarreta também a solução para um problema normativo (saber o que conta como justificado) (MIKHAIL, 2011, p. 27-32). E, esse é o ponto decisivo, a solução para o problema descritivo acarreta uma

[1] Para a lista de restrições para o que conta como juízos ponderados, ver Rawls (1951, p. 181-183; para a lista de restrições para o que conta como um juíz moral competente, ver Rawls (1951, p. 178-180).

solução para o problema normativo porque o que é descrito não são fatos brutos ou o simples desempenho moral de indivíduos comuns, mas uma capacidade ou competência confiável de distinguir entre o certo e o errado de um juiz moral competente.

O ponto importante desse procedimento para a leitura de Mikhail é a sua caracterização enquanto uma investigação empírica. Claro, a tese de que a solução para o problema descritivo acarreta a solução para um problema normativoé o típico problema filosófico a ser resolvido "a partir da poltrona", sem recurso às ciências empíricas. Mas a outra parte do projeto, de encontrar um conjunto de princípios que explicaria os juízos morais ponderados de um juíz moral competente, é, dado tudo o que Rawls diz, o tipo de investigação que situa a teoria moral como uma ciência empírica. Embora os princípios de uma explicação possam ser formulados a partir da poltrona, como Rawls esboça no *Outline*, eles são apresentados como hipóteses que devem ser confirmadas empiricamente.

Mikhail acredita que uma análise das seções 4 e 9 de *TJ,* assim como do artigo de 1975, "The Independence of Moral Theory"(*Independece*, daqui em diante), confirma que Rawls não abandonou essa concepção em certo sentido "naturalista"de teoria moral. A analogia gramatical presente na seção 9 compõe o núcleo da interpretação defendida por Mikhail. Ao utilizar essa analogia Rawls estaria propondo que o objetivo principal da teoria moral édescrever um objeto factual que Mikhail caracteriza como "o sistema moral da mente/cérebro humana"(I-Morality). Esse sistema moral seria um composto por certas regras ou princípios operativos que estariam implicados no uso de conceitos morais pré-teóricos em nossos juízos morais ponderados, princípios que regem a nossa competência de distinguir entre o certo e o errado (MIKHAIL, 2011, p. 63). Basicamente, Rawls teria o mesmo programa empírico dos linguistas que seguem a tradição de Chomsky, exceto por buscar descrever uma capacidade moral em vez de uma capacidade linguística. Mikhail não acredita que Rawls tenha desenvolvido esse programa, mas defende que ele teria esboçado as suas bases.

Uma observação preliminar sobre a interpretação oferecida por Mikhail éque ela se concentra, exclusivamente, no *Outline,* nas seções 4 e 9 de *TJ* e no *Independence*. Mikhail mesmo reconhece que depois de 1975, data de publicação do *Independence*, Rawls talvez tenha abandonado o seu programa linguístico e o substituído por um projeto construtivista (MIKHAIL, 2011, p. 267, 274, 294-295). Mikhail acreditaa que a sua interpretação éa melhor de *TJ,* mas ele não a apresenta como a melhor interpretação dos textos escritos depois de 1975. Na seção seguinte eu defenderei que a interpretação de Mikhail éinadequada mesmo como uma leitura de *TJ* e do *Independence,* embora ela seja, em linhas gerais, apropriada como uma caracterização do projeto do *Outline.* Mas antes nós precisamos avaliar os argumentos de Mikhail.

A concepção de teoria moral em *TJ* parte do pressuposto de que cada pessoa além de uma certa idade e com uma capacidade intelectual mínima desenvolve sob circunstâncias sociais normais um senso de justiça. Esse senso de justiça écompreendido como uma capacidade moral que explicaria como podemos proferir um potencialmente infinito número de juízos morais nas mais variadas circunstâncias (RAWLS, 1999, p. 41) - assim como para Chomsky uma competência linguística explicaria como uma criança, após ter adquirido um certo domínio da sua linguagem, seria capaz de fazer intuitivamente uma série de juízos sobre a gramaticalidade de variadas sentenças (MIKHAIL, 2011, p. 4). Para Rawls a teoria moral pode ser vista "provisoriamente"como a"tentativa de descrever nossa capacidade moral", "senso de justiça"ou "sensibilidade moral", ou, como Mikhail propõe, uma investigação sobre a nossa competência moral. Para exemplificar esse projeto Rawls elabora uma analogia com a teoria linguística de Chomsky, em que esclarece que descrever um senso de justiça écomo a tarefa de descrever um senso de gramaticalidade: assim como um gramático tenta caracterizar a habilidade dos nativos de reconhecer sentenças corretamente compostas formulando um conjunto de princípios que façam as mesmas discriminações que o falante nativo, o teórico moral teria a tarefa de caracterizar uma capacidade moral formulando princípios que sistematize as regras implicitamente utilizadas pelas pessoas nos seus juízos morais ponderados (RAWLS, 1999, p. 41).

Embora esse projeto não seja mais descrito como um programa empírico como no *Outline*, a teoria desenvolvida em *TJ* édescrita como uma "teoria dos sentimentos morais"que esquematiza os princípios que "governam nossas capacidades morais". No *Independence* Rawls classifica a teoria moral, assim como o equilíbrio reflexivo, como um "tipo de psicologia", e descreve o teórico moral "como um observador, por assim dizer, que busca delinear a estrutura das atitudes e concepções morais de outras pessoas."(RAWLS, 1975, p. 7-9). Esse tipo de referência é coletada por Mikhail em defesa de sua interpretação. Mas o que nós precisamos analisar é a ideia do equilíbrio reflexivo, pois écom essa noção que Rawls pretende explicar o que significa afirmar que uma teoria moral "descreve"um senso de justiça.

Em *TJ* Rawls discute o equilíbrio reflexivo nas seções 4 e 9. Nós sabemos que para Rawls os princípios de justiça são justificados porque eles seriam escolhidos na posição original, mas que isso não resolve o problema da justificação porque a posição original écomposta por um conjunto de restrições morais que precisam ser elas mesmas justificadas (RAWLS, 1999, p. 17). O equilíbrio reflexivo éintroduzido na seção 4 para justificar a própria posição original. Cito a passagem:

> In searching for the most favored description of this situation we work from both ends. We begin by describing it so that it represents generally shared and preferably weak conditions. We then see if these conditions

> are strong enough to yield a significant set of principles. If not, we look for further premises equally reasonable. But if so, and these principles match our considered convictions of justice, then so far well and good. But presumably there will be discrepancies. In this case we have a choice. We can either modify the account of the initial situation or we can revise our existing judgments, for even the judgments we take provisionally as fixed points are liable to revision. By going back and forth, sometimes altering the conditions of the contractual circumstances, at others withdrawing our judgments and conforming them to principle, I assume that eventually we shall find a description of the initial situation that both expresses reasonable conditions and yields principles which match our considered judgments duly pruned and adjusted. This state of affairs I refer to as reflective equilibrium. It is an equilibrium because at last our principles and judgments coincide; and it is reflective since we know to what principles our judgments conform and the premises of their derivation. (RAWLS, 1999, p. 18).

Rawls defende que a posição original éjustificada porque ela émoldada de tal modo que os princípios de justiça que dela resultam estão de acordo com os nossos juízos morais ponderados em equilíbrio reflexivo[2]. Mikhail defende que nessa passagem o equilíbrio reflexivo éoficialmente definido como um "estado de coisas", a saber, um estado no qual háuma relação de coerência (equilíbrio) entre juízos morais ponderados, princípios de justiça e uma teoria moral, e não éum método, uma técnica ou um procedimento para ser empregado. Mikhail oferece a sua própria definição: um "estado de coisas alcançado quando o *teórico moral* sabe os princípios com os quais o conjunto de juízos morais ponderados se conformam, e as premissas daqueles princípios." (MIKHAIL, 2011, p. 205).

Em *Independence* Rawls define o equilíbrio reflexivo empregado em *TJ* como *amplo* (*wide*), justamente por ser um equilíbrio entre juízos morais ponderados, princípios de justiça e outras descrições alternativas possíveis desse senso de justiça. Ele defende que "adotando o papel de teóricos morais observadores, nós investigamos os princípios que as pessoas reconheceriam"se elas tivessem a "oportunidade de considerar outras concepções plausíveis"(RAWLS, 1975, p. 8). No *Outline* não hámenção para essa consideração de outras possibilidades de descrição, o que implica que a explicação oferecida no *Outline* descreve o senso de justiça de alguém mais ou menos como ele é, enquanto que a descrição oferecida pelo equilíbrio reflexivo amplo em *TJ* pode, provavelmente irá, oferecer uma descrição que requer drásticas revisões em um senso de justiça[3]. Conforme interpreta Mikhail, ao invocar a categoria de equilíbrio reflexivo amplo Rawls concede a possibilidade de que quando se édada

[2] Rawls descarta a categoria de juízes morais competentes apresentada no *Outline*, mas mantém a categoria de juízos morais ponderados mais ou menos inalterada.

[3] Por essa razão écomum definir o equilíbrio reflexivo utilizado no *Outline* como estreito (sobre a distinção entre as versões estreita e ampla, ver DANIELS 1996, p. 66-72).

a uma pessoa a oportunidade para ela refletir sobre a "teoria empiricamente adequada do seu senso de justiça", essa oportunidade de reflexão pode alterar dramaticamente o seu sistema de crenças (MIKHAIL, 2010, p. 23). O *rationale* para essa ampliação do equilíbrio reflexivo, Rawls indica, éevitar que a teoria moral seja acusada de "conservadora." (RAWLS, 1975, p. 7-8). Mikhail explica que o *rationale* se deve ao fato de que Rawls apresenta a sua teoria de justiça como uma teoria preferível a teorias rivais, e espera mostrar aos seus leitores com inclinações utilitaristas que eles talvez estejam considerando inadequadamente as consequências dos seus juízos morais ponderados (MIKHAIL, 2011, p. 24).

A interpretação de Mikhail éque tanto no *Outline* quanto em *TJ* Rawls apresenta o equilíbrio reflexivo como um "estado de coisas alcançado no curso de avaliar descrições alternativas do senso de justiça." (MIKHAIL, 2010, p. 17). Embora Habermas tenha aludido a essa interpretação (HABERMAS, 1995, p. 120), Mikhail épara o meu conhecimento o único autor a defendê-la extensivamente. Na seção seguinte eu gostaria de destacar alguns problemas com essa leitura. Eu defenderei que nós deveríamos interpretar o equilíbrio reflexivo (sempre amplo, daqui em diante) não como um estado de coisas, mas como um processo de reflexão que o *agente moral* deve empreender.

Equilíbrio reflexivo como um método de reflexão

A intepretação que defenderei éesta: o equilíbrio reflexivo éum método, e o estado de coisas constituído por uma relação de ajuste mútuo entre juízos, princípios e teorias morais éo resultado *do uso,* por parte do agente moral, desse método. A minha interpretação se distingue da de Mikhail por defender não apenas que o equilíbrio reflexivo éum método, mas por defender que ele éum método que deve ser empregado *pelo* agente moral, não pelo teórico moral, para refletir sobre a melhor descrição do seu senso de justiça. Essa interpretação não éoriginal e é, eu diria, a interpretação padrão do equilíbrio reflexivo, defendida por autores como Daniels (1996), De Paul (1993) e Scanlon (1992, 2003, 2014). Mikhail acredita que essa interpretação padrão não éfiel ao texto de *TJ* e *Independence.* Nesta seção eu argumentarei que ela é.

Considere esta passagem que pode ser encontrada no *Independence,* os itálicos são meus*:*

> *The procedure of reflective equilibrium* does not, by itself, exclude this possibility, however unlikely it may be. For in the course of achieving *this state*, it is possible that first principles should be formulated that seem so compelling that they lead us to revise all previous and subsequent judgments inconsistent with them. *Reflective equilibrium requires only that the agent makes these revisions with conviction and confidence*, and continues to affirm these principles when it comes to accepting their consequences in practice. (RAWLS, 1975, p. 8).

Esta passagem, eu penso, representa um embaraço àleitura oferecida por Mikhail. Rawls afirma que o equilíbrio reflexivo *requer* que o agente faça certas revisões no seu sistema de crenças. Como Mikhail pode explicar que o equilíbrio reflexivo faça requisiçõesao agente? Talvez ele poderia argumentar que se se demonstra ao agente que as suas crenças constituem um estado de coisas incoerente ou inconsistente, então esse *fato* ou *esse* estado de coisas *requer* que o agente faça alguma coisa - a saber, revise os seus juízos morais ponderados. E fatos podem realmente exigir revisões no nosso sistema de crenças, como quando dizemos que o fato de que agora faz frio *requer* de mim que eu rejeite a minha crença de que eu não preciso de um agasalho, ou fato de que o *Titanic* afundou *requereu* das pessoas da época que abandonassem a crença de que ele era um navio inaufragável. Mas Rawls não estáapenas dizendo que o equilíbrio reflexivo requer que o agente *mude* de crenças, ele diz que ele requer que essas revisões sejam feitas *de uma certa maneira* - "com convicção e confiança e que continue a aceitar esses princípios quando tiver de aceitar as suas consequências na prática". Não acredito que Mikhail pode explicar em que sentido o equilíbrio reflexivo *requer um modo de* fazer revisões.

Eis a minha sugestão: háuma certa ambiguidade no uso do nome "equilíbrio reflexivo". Em muitos casos o termo éusado para descrever um certo estado de coisas–um estado tal em que juízos, princípios e teorias morais estão em uma relação de ajuste mútuo. É isso que Rawls tem em mente quando ele afirma que a sua concepção de justiça descreve o que alguém afirmaria se estivesse em equilíbrio reflexivo. Mas em sentenças como "o equilíbrio reflexivo requer apenas que o agente faça essas revisões com convicção e confiança"a ideia de equilíbrio reflexivo não figura como uma descrição de um estado de coisas, mas, ao invés, aponta para uma concepção de *como* alguém deve proceder para decidir da melhor maneira possível que teoria da justiça aceitar e o que acreditar. O *processo* pelo qual se chega ao estado de equilíbrio reflexivo (isto é, encontrar juízos morais ponderados, formular princípios que expliquem esses juízos, e revisar os juízos, os princípios ou ambos em caso de conflito) essse processo em *si,* que em *TJ* Rawls chama de "curso hipotético de reflexão", éo próprio equilíbrio reflexivo (RAWLS, 1999, p. 18). Nós podemos dizer que o *estado de equilíbrio reflexivo* éalcançado *como resultado* do conscienciosoe adequado *uso do* método do equilíbrio reflexivo. Assim, a hipótese defendida em *TJ* éa de que se uma pessoa *seguir apropriadamente o curso hipotético de reflexão definido pelo equilíbrio reflexivo,* então ela aceitaráa justiça como equidade como a teoria mais razoável, dadas as suas convicções morais mais firmes (RAWLS, 1999, p. 17-18).

A "descrição"que a teoria moral fornece édescrita como "provisória"e Rawls insiste em destacar essa característica (RAWLS, 1999, p. 41). O que ele quer dizer com provisória? Mikhail interpreta que a descrição éprovisória porque o teórico moral estáaberto àpossibilidade de que a descrição ofere-

cida pode sofrer alterações no curso do processo de se atingir equilíbrio reflexivo (MIKHAIL, 2011, p. 293). O que essa explicação do Mikhail mostra éque o conteúdo de uma descrição oferecida éprovisório, mas Rawls utiliza o "provisório"como um adjetivo para a natureza da teoria moral em si. Rawls estádizendo que a teoria é*provisoriamente* descritiva, e não que o que ela descreve éprovisório. Se o conteúdo de uma descrição pode alterar o seu objeto de estudo, necessitando ser readequado àluz dessa alteração, como Mikhail propõe, ela ainda é*definitivamente* uma teoria descritiva. O que Rawls estádizendo éque uma vez que a teoria exige revisões no senso de justiça, ela jádeixa de ser uma descrição[4].

Rawls afirma que a teoria da justiça descreve nosso senso de justiça *em* equilíbrio reflexivo. De acordo com o que foi anteriormente exposto, isso significa que a teoria moral descreve o senso de justiça que a pessoa *teria* caso ela seguisse o método do equilíbrio reflexivo. Seguir o método significa assumir uma postura crítico-reflexiva diante de seus juízos morais ponderados àluz da descrição desse senso de justiça que o teórico moral oferece. Em larga medida a descrição do teórico moral serácontraditória com muitos dos juízos morais ponderados pré-reflexão, mas se a descrição for adequada, e o agente moral for razoável, o agente poderáver que a descrição oferecida éa descrição daquilo que ele realmente sustenta em questões de justiça, e esse reconhecimento levaráo agente a revisar o seu sistema de crenças. A teoria moral não descreve fatos naturais, mas descreve o *que alguém aceitaria em questões de justiça se fosse apropriadamente reflexivo,* ou se estivesse em equilíbrio reflexivo[5]. Na medida em que o equilíbrio reflexivo, entendido como um método, depende de certas assunções normativas sobre o que conta como uma deliberação *apropriada* ou *correta,* então o objeto da descrição éum conjunto de crenças e princípios que estão de acordo com essas restrições normativas (RAWLS, 1975, p. 8). O objeto da descrição não énenhum fato empírico observável. Mikhail replicou a essa linha de raciocínio argumentando que ela ignora que para Rawls a solução para um problema descritivo (descrever uma compe-

[4] Scanlon distingue entre uma interpretação deliberativa e uma intepretação descritiva do equilíbrio reflexivo. De acordo com a primeira, que éa que estou defendendo nesta seção, o equilíbrio reflexivo éum método que um indivíduo deve adotar para descobrir o que acreditar sobre questões de justiça. De acordo com segunda, que éa leitura favorecida por Mikhail, o objetivo do método seria caracterizar uma concepção de justiça sustentada por uma pessoa ou grupo (SCANLON, 2003, p. 142). Scanlon defende uma leitura deliberativa com o argumento de que ela faz mais sentido diante do modo como o processo de revisibilidade écaracterizado por Rawls.

[5] Essa étambém a conclusão de Daniels. Ele defende que nós deveríamos recusar a analogia com a linguística para ilustrar a natureza da teoria moral por duas razões. Primeiro, o *rationale* para a revisabilidade inerente ao equilíbrio reflexivo não éo de corrigir aqueles juízos morais ponderados que não refletem a real competência moral do indivíduo, como éna linguística, mas, sim, éo de formular um senso de justiça que nós, como pessoas, queremos ver realizado. Segundo, com o equilíbrio reflexivo Rawls não estábuscando descrever nenhuma competência real, mas, ao invés, estábuscando articular uma competência*ideal,* isto é, uma competência que a pessoa teria se ela fosse persuadida por argumentos filosóficos e revisasse o seu sistema de crença de acordo com esses argumentos. (DANIELS, 1996, p. 71-72).

tência moral) étambém a solução para o problema normativo (descobrir princípios de justiça justificados), e que ambos os problemas estão interconectados (MIKHAIL, 2011, p. 94). Essa réplica, porém, erra o alvo: o que a interpretação sugerida afirma éque não háum problema descritivo a ser resolvido em *TJ.*

Alguém poderia objetar: mas se a teoria moral éprovisioriamente descritiva, ela ainda é, em algum momento, descritiva. A minha resposta éesta: quando o teórico moral apresenta um conjunto de juízos e princípios que alguém deve aceitar se quiser aceitar juízos e princípios justificados, ele tem de começar de algum lugar. Ele analisa juízos morais ponderados que são amplamente compartilhados e sustentados com confiança e convicção, como os juízos de repúdio àescravidão e de tolerância religiosa, e não encontra razões para duvidar da razoabilidade desses juízos. Ele formula então uma teoria que a seu ver articula os vários conceitos, ideias e princípios afirmados nesses juízos. O objetivo éprovisório porque uma vez que o teórico analisa o comportamento real dos indivíduos, e a totalidade dos juízos que eles realizam, ele percebe que a o que ele oferece estámais para uma prescrição do que para uma descrição: o conteúdo da sua teoria representa o que ele espera que ninguém teria razões para pensar que éirrazoável após consideração ou reflexão racional, e o que alguém, que estivesse disposto a revisar o seu senso de justiça e fosse razoável, aceitaria como correto em questões de justiça[6].

A autoridade do teórico moral

Háuma diferença fundamental entre a interpretação de Mikhail e a que estou sugerindo: de acordo com a minha, mas não com a de Mikhail, o que o agente moral pode reconhecer como aceitável após reflexão apropriada é*determinante* para estipular o que conta como justificado. De acordo com a posição de Mikhail, aceitabilidade em reflexão éirrelevante: o teórico moral tem a palavra última sobre qual éa concepção de justiça justificada. Nesta seção pretendo mostrar por que Mikhail, mas não Rawls, estácomprometido com essa posição.

Mikhail identifica que háuma distinção entre dois tipos de regras ou princípios. Háos princípios expressos, "enunciados que uma pessoa verbaliza na tentativa de descrever, explicar ou justificar os seus juízos", que não são relevantes para o teórico moral, e háos princípios denominados de operativos, que são aqueles"realmente operativos no exercício do senso de justiça" que são identificados por investigação empírica conduzida pelo teórico moral (MIKHAIL, 2011, p. 19-21). Rawls afirma que uma descrição adequada do senso de justiça de uma pessoa envolve princípios e construções teóricas "que vão muito além das normas e padrões citados na vida diária." (RAWLS, 1999, p. 41-42). Dessa

[6] Para uma compreensão de teoria moral nessas linhas ver Scanlon (1992).

afirmação Mikhail interpreta que Rawls estáafirmando tanto que um indivíduo pode reportar falsamente qual princípio regula a sua competência moral de distinguir o certo e o errado, quanto estádefendendo que o teórico moral não assume que a pessoa "pode tornar-se consciente dessas regras por meio da introspecção". "Como resultado de uma investigação empírica", essas regras ou princípios operativos estão "além de qualquer conscienciosidade real ou potencial." (MIKHAIL, 2011, p. 19, 50-51). Mikhail defende que uma sentença é*gramatical* se ela estáde acordo com princípios operativos, e ela é*aceitável* se ela é, para o falante nativo, natural ou conforme às suas intuições linguísticas. Se a descrição oferecida pelo teórico moral não for aceitável após reflexão apropriada para o agente cujo senso de justiça estásendo descrito, esse fato por si sónão éuma razão para o teórico alterar a sua descrição: éo teórico moral quem estáequipado para fornecer uma descrição de princípios operativos. O acesso introspetivo e reflexivo em primeira pessoa estásujeito a distorções que uma investigação empírica, realizada em terceira pessoa por um observador externo, não está (MIKHAIL, 2011, p. 286).

A passagem de Rawls citada algumas linhas atrás não autoriza em nenhum momento essa assimetria forte que determina que uma sentença gramatical não é(ou não pode ser) acessível via introspecção ou reflexão[7]. Não se segue da afirmação de que uma descrição do senso de justiça envolve teorização e necessita de recursos que vão além do que écitado na vida diária que esses recursos não são nem mesmo potencialmente acessíveis àconsciência via introspecção e reflexão. Mas é*nessa* afirmação que Mikhail apoia a sua tese interpretativa (MIKHAIL, 2011, p. 235). Embora um indivíduo possa estar equivocadosobre quais princípios de fato regulam o seu senso de justiça, ou acreditar que um juízo é*aceitável* quando ele não égramatical, ou acreditar que ele éinaceitável quando ele estáde acordo com regras válidas, Rawls mantém, e de fato enfatiza, que o que éjustificado ou válido tem de poder ser visto como aceitável em, ou após, reflexão apropriada. Longe de buscar um distanciamento das intuições morais ordinárias (no sentido especificado por Mikhail), Rawls sustenta que um princípio válido éum princípio que se aplicado "nos levaria a fazer os mesmos juízos sobre a estrutura básica da sociedade que nós agora fazemos *intuitivamente* e nos quais nós temos a maior *confiança*", e que em momentos de dúvida e hesitação "oferecem uma resolução que nós podemos endossar em reflexão"[8]. Ele também escreve que nós teríamos algum interesse em seguir os princípios escolhidos na posição original, considerando que essa escolha éapenas hipotética e não factual, porque a posição original

[7] Mesmo no Outline, artigo em que a teoria moral étratada como uma investigação empírica, Rawls não defende que os princípios podem ser inacessíveis ao agente. Pelo contrário, ele salienta que eles *têm de ser acessíveis* para que possam ser vistos como justificados (RAWLS, 1951, p. 188).

[8] "In cases where our present judgments are in doubt and given with hesitations, these principles offer a resolution which we can affirm on reflection."(RAWLS, 1999, p. 17).

incorpora condições que nós de fato aceitamos, ou, se não aceitamos, que poderíamos ser persuadidos a aceitar por reflexão filosófica (RAWLS, 1999, p. 19). A linguagem do equilíbrio reflexivo não apenas rejeita uma distinção entre gramaticalidade e aceitabilidade ("intuitivamente apelante", "sua confiança", "ela pode aceitar"), como parece definir a primeira a partir da segunda.

Eu gostaria de encerrar indicando como a distinção entre princípios operativos e expressos, e gramaticalidade e aceitabilidade, produz uma imagem de teoria moral que Rawls não aceitaria. Se o teórico moral éoúnico que estáem posição de saber quais são os princípios operativos que realmente descrevem o efetivo senso de justiça, se esses princípios são potencialmente inacessíveis àreflexão de agentes morais reais cujo senso de justiça eles descrevem, e se esses princípios são *justificados* em virtude dessa adequação descritiva, então o problema de descobrir princípios morais justificados passa a ser algo sobre o qual apenas teóricos morais estão *capacitados* a dar um veredito.

O equilíbrio reflexivo parece justamente enfatizar que aqueles que são capazes de submeter os seus juízos morais ponderados ao crivo de uma reflexão "ampla"a partir de teorias morais, seus pressupostos e argumentos filosóficos correspondentes, estão em melhor posição para pensar sobre questões de justiça justificadamente. Teóricos morais são assim naturalmente uma classe de pessoas em boa posição para propor o que éo justo e o que devemos fazer. Mas disso não se segue que teóricos morais possuem a *autoridade* do veredito. Na seção 87 de *TJ*, Rawls defende que justificar um princípio não émostrar que ele éverdadeiro ou falso, mas éuma questão de endereçar argumentos àqueles que discordam de nós, ou a nós mesmos quando estamos em dúvida sobre o que aceitar. O objetivo da justificação é*prático* na medida em que justificação é"conciliação através da razão", éum argumento endereçado ao outro a partir de premissas que *ambas as partes aceitam* (RAWLS, 1999, p. 508). Para Rawls o problema da justificação, e de encontrar quais princípios são justificados, éum problema que envolve as partes concernidas no conflito de interesses (no caso, a sociedade). Mikhail não analisa a seção 87.

A imagem de teoria moral que atribuo a Rawls éradicalmente diferente daquela proposta por Mikhail. Mikhail descreve um modelo de justificação e de teoria em que a solução para os problemas normativos *vemde fora* dessa prática argumentativa, como se a solução fosse *anunciada* pela investigação do teórico. No modelo efetivo de Rawls, contudo, a solução me parece *vir de dentro* dessa prática. O papel da teoria moral éo de servir como um elemento *qualificador* dessa prática argumentativa: o acesso a teorias morais permite àcomunidade compreender melhor as implicações dos seus juízos morais ponderados, as interconexões entre os problemas que ela enfrenta e os comprometimentos das possíveis soluções. Antes que o anunciador de soluções para os problemas normativos, nessa imagem de teoria moral o teórico moral éum *facilitador* do processo a partir do qual essas soluções emergem. Eu acre-

dito que esse modelo de teoria moral émuito claramente afirmado por Rawls nos seus últimos escritos, mas nesta seção eu pretendi mostar por que ele jáestápresente em *TJ.*

Considerações finais

Talvez os esforços interpretativos de Mikhail possam ser justamente resumidos como uma tentativa de defender que Rawls mantém em *TJ* a mesma estratégia do *Outline:* encontrar uma solução para o problema normativo (quais princípios são justificados?) através da investigação de um problema unicamente descritivo (quais princípios regem a nossa capacidade moral?). Eu acredito que essa intepretação éequivocada. Propus queem *TJ* o equilíbrio reflexivo funciona como um método que estipula uma certa concepção normativa do que édeliberar ou refletir *adequadamente.* Uma vez que essa interpretação do método éaceita, como eu argumentei que deveríamos, a afirmação de que uma teoria moral descreve princípios de justiça *em* equilíbrio reflexivo passa simplesmente a significar que uma teoria moral "descreve"o que nós aceitaríamos como a melhor descrição do nosso senso de justiça, se estivéssemos refletindo *adequadamente,* ou o que nos parece mais razoável após reflexão adequada.

Na seção passada eu indiquei uma consequência a meu ver importante dessa reformulação do equilíbrio reflexivo, a saber, que ela evita que a solução para os problemas normativos que uma comunidade enfrenta *venhade fora* da prática argumentativa dessa comunidade. Eu gostaria agora de encerrar destacando outra implicação de se aceitar que o problema normativo não éum problema descritivo. Rawls baseia a sua confiança de que nós aceitaríamos a sua teoria em equilíbrio reflexivo porque ele acredita que ela articula as ideias, preceitos e razões que explicam certos juízos morais ponderados que *de fato* sustentamos com confiança e convicção, como os juízos de repúdio àescravidão e de tolerância religiosa. Isso significa, devemos conceder a Mikhail, que ele acredita que a sua teoria tem um certo mérito descritivo. Mas que a teoria reflete esses juízos morais ponderados que são de fato sustentados, e que nós a aceitaríamos em equilíbrio reflexivo, não parece ser o que importa para Rawls. O que importa éque após consideração racional nós não temos razões para desconfiar que a teoria, ou esses juízos nos quais ela estábaseada, éirrazoável.

Referências bibliográficas

DANIELS, N. *Justice and Justification:* Reflective Equilibrium in Theory and Practice. Cambridge: Cambridge University Press, 1996.

DEPAUL, M. *Balance and Refinement:* Beyond Coherence Methods of Moral Inquiry. London: Routledge, 1993.

HABERMAS, J. Reconciliation through the public use of reason: Remarks on John Rawls'Political Liberalism. *The Journal of Philosophy*, v. 92, n. 3, 1995, p. 109-131.

MIKHAIL, J. *Elements of Moral Cognition:* Rawls' Linguistic Analogy and the Cognitive Science of Moral and Legal Judgment. Cambridge: Cambridge University Press, 2011.

_____. Rawls' Concept of Reflective Equilibrium and its Original Function in 'A Theory of Justice'. *Washington University Jurisprudence Review*, v. 3, n. 1, p. 1-30, 2010.

RAWLS, J. *A Theory of Justice.* Revised Edition. Cambridge: Harvard University Press, 1999.

_____. *A Theory of Justice.* Original Edition. Cambridge: Harvard University Press, 1971.

_____. The Independence of Moral Theory. *Proceedings and Addresses of the American Philosophical Association*, v. 48, p. 5-22, 1975.

_____. Kantian Constructivism in Moral Theory. *The Philosophical Journal*, v. 77, n. 9, p. 515-572, 1980.

_____. Outline of a Decision Procedure for Ethics. *The Philosophical Review*, v. 60, n. 2, p. 177-197, 1951.

SCANLON, T. *Being Realistic about Reasons*. Oxford: Oxford University Press, 2014.

_____. "Rawls on Justification". In: FREEMAN, S. (Ed.). *The Cambridge Companion to Rawls*. Cambridge: Cambridge University Press, 2003, p. 139-168.

_____. The Aim and Authority of Moral Theory. *Oxford Journal of Legal Studies*, v. 12, n. 1, p. 1-23, 1992.

ARGUMENTOS
Revista de Filosofia

Santiago Prono*

Democracia deliberativa e ideal de reciprocidad. Un análisis desde la teoria del discurso

RESUMEN

El objetivo del presente trabajo es mostrar que el ideal deliberativo de reciprocidad, introducido por Rawls en sus escritos sobre el liberalismo político para la definición de cuestiones de justicia básica, resulta conceptualmente inviable en el marco teórico de la democracia deliberativa de Habermas. La razón ofrecida como fundamento, sostiene que dicho ideal no sólo implica un uso estratégico de la racionalidad, que como tal deslegitima el procedimiento intersubjetivo y público de deliberación racional para la fundamentación de decisiones y normas que establece el concepto de política deliberativa del filósofo alemán, sino que también entra en contradicción con el principio del discurso argumentativo, el cual define las condiciones de validez de las decisiones adoptadas en tal procedimiento al asegurar su valor epistémico.

Palabras-clave: Habermas; Rawls; Reciprocidad; Democracia Deliberativa; Discurso.

ABSTRACT

The aim of this paper is to show that the deliberative ideal of reciprocity, introduced by Rawls in his writings on political liberalism to define matters of basic justice, it is not conceptually feasible in the theoretical framework of Habermas's deliberative democracy. The reason offered as foundation, argues that this ideal not only involves one strategic use of rationality, which discredits the intersubjective and public procedure of rational deliberation for the foundation of decisions and norms that establishes the concept of deliberative politics of the German philosopher, but it also contradicts the principle of argumentative discourse, which defines the conditions of validity of the decisions adopted in a such procedure because it ensures his epistemic value.

Keywords: Habermas; Rawls; Reciprocity; Deliberative Democracy; Discourse.

* Dr. en Filosofía, investigador del Consejo Nacional de Investigaciones Científicas y Técnicas (Argentina), Docente-Investigador de la Universidad Nacional del Litoral (Santa Fe, Argentina).

Introducción

El ideal de reciprocidad adopta un lugar fundamental en la concepción deliberativa de la democracia sostenida por J. Rawls en sus últimos escritos. Este ideal exige que los ciudadanos deben abstenerse de apelar a sus creencias comprehensivas (de tipo moral, filosófico, o religioso) cuando participan en el procedimiento decisorio sobre temas de justicia básica que regulan el ordenamiento institucional del estado de derecho. El argumento del autor para justificar esta restricción, es que tales creencias, aun cuando quienes las sostienen piensen que están correctamente fundamentadas, impiden alcanzar consensos ya que se relacionan con el tipo de valores sobre los que, en general, no sólo no hay acuerdo, sino que además tampoco se admite la posibilidad de revisión cuando entran en juego posturas antitéticas. De este modo, y a fin de intentar alcanzar consensos en el momento clave de decidir, tales ciudadanos tienen que actuar sobre la base de un cálculo que implica la previa exclusión de aquellas creencias comprehensivas[1].

Muchos son los autores que en el ámbito de la Filosofía política contemporánea analizan este tema, tanto desde el punto de vista exclusivo del liberalismo político (BUCHANAN 1990, BARRY 1995, NUSSBAUM 2009, GAUS 2010, NEUFELD 2010, LISTER 2011, GARRETA LECLERQ, 2009a, 2009b, 2012.), como así también teniendo en cuenta un concepto de política cercano al asumido por la democracia deliberativa (COHEN 1997, 2011, MARTINS 2008, RICO MOTOS 2009, ROBLES 2011). Ahora bien, considerado este ideal desde la concepción habermasiana de esta teoría política, el mismo comporta una serie de implicancias que resultan conceptualmente problemáticas de articular con los presupuestos filosóficos sobre los que esta se fundamenta. El punto en cuestión, es que el ideal deliberativo de Rawls expresa un carácter contractual que por su correspondiente sentido pragmático-estratégico se contrapone a las pretensiones universales de validez del concepto de política deliberativa sostenido por el filósofo alemán, basado en el concepto de racionalidad comunicativa y en el principio del discurso argumentativo, lo cual afecta el valor epistémico de las decisiones adoptadas en el marco del procedimiento intersubjetivo de deliberación racional en el que dicho principio tiene vigencia.

A fin de justificar esta tesis, el presente trabajo analiza el tema señalado en base a la siguiente estructura expositiva. Luego de una presentación general de la democracia deliberativa de Habermas en la que se explicitan algunos de sus fundamentos teóricos que posteriormente se analizan (I), se expone el

[1] Desde el comienzo, y luego volveremos sobre esto, es necesario aclarar que Rawls admite que doctrinas comprehensivas pueden formar parte del proceso decisorio, pero sostiene que en el momento de decidir ellas deben ser excluidas. La idea es que en la decisión los interlocutores involucrados den una justificación política de sus propuestas, una justificación que de ningún modo dependa de la aceptación de sus doctrinas comprehensivas (RAWLS, 1999b: 142 ss.; 1999c: 197 ss.).

ideal deliberativo de reciprocidad que sostiene Rawls, junto con la lectura del mismo que conciben algunos de sus comentaristas (II). A continuación se analiza la teoría de la acción comunicativa (III) y el principio del discurso argumentativo (IV), que constituyen el trasfondo teórico de la filosofía política habermasiana. Esto permitirá exponer, y analizar, los problemas conceptuales que resultan inherentes al hecho de concebir el ideal rawlsiano de reciprocidad en la democracia deliberativa de Habermas. Las conclusiones generales sólo estriban en un breve resumen de los resultados alcanzados en función de los argumentos presentados (V).

Democracia deliberativa. Una (muy breve) introducción

La idea de democracia implica, a la vez, un concepto descriptivo y normativo. Es descriptivo porque da cuenta del modo en que se toman las decisiones en un estado democrático, por ejemplo en el Parlamento donde se discute sobre determinadas cuestiones; pero también, y fundamentalmente, es normativo porque exige que las decisiones del gobierno sean el resultado de la participación, directa o indirecta, de los ciudadanos como uno de los criterios de su legitimación política. Ahora bien, el concepto habermasiano de la democracia deliberativa también implica deliberación, intercambio de opiniones antitéticas que, al menos en principio, tiene que confrontarse en términos de argumentos para intentar llegar a la mejor decisión posible y así lograr acuerdos racionalmente motivados. En este contexto, el principio básico de la democracia deliberativa no es el principio de la mayoría, sino (como veremos) el *principio del discurso argumentativo*. Así, la teoría de la democracia deliberativa pretende constituirse en criterio de justificación de la validez de las decisiones políticas, y de la consolidación de los sistemas democráticos del estado de derecho. Naturalmente, esta legitimación de la práctica democrática se presenta como un objetivo al que deberíamos tender para dirimir las diversas pretensiones de validez que en tal contexto se presenten, *sin* por ello concebir que pueda efectivamente alcanzarse en todos los casos un consenso[2].

La teoría de la democracia deliberativa también pretende articular el desempeño de las instituciones formales del estado de derecho, que constituyen el contexto en el que se justifican y toman decisiones, con los aportes de las organizaciones de la sociedad civil, que ocupan un lugar preponderante con

[2] En Habermas es necesario no confundir el uso del término "deber", o "tener que" (*müssen*), con el "deber" en el sentido del verbo alemán *sollen*, que comporta un sentido moral, porque con el uso de aquellos términos el autor sólo pretende dar cuenta de una necesidad lógica o pragmática explicitada a partir de la reconstrucción de los presupuestos operantes en la formulación de argumentos mediante el uso comunicativo del lenguaje (HABERMAS, 1995pp. 114 ss.; 1994, p. 19, 61 ss., y 399 ss.; al respecto véase también De ZAN, 2004, p. 59 ss.).

pretensiones de influir en el ámbito de la política democrática, en el sentido de que constituyen el contexto de descubrimiento de los temas y problemas que afectan a la sociedad global, que tienen que ser analizados por aquellas instituciones formales de la democracia. Así, esta teoría establece una conexión entre los espacios públicos formales e informales de la política. Se trata, sin embargo, de una conexión que se realiza adoptando un posicionamiento equidistante entre el liberalismo y el republicanismo, y por el cual la democracia deliberativa incorpora algunos principios de cada uno y se diferencia de otros para integrarlos de una forma nueva y original en base a sus propios presupuestos filosóficos, por lo cual no se trata de una mera combinación o síntesis entre ambas tradiciones de la política (HABERMAS, 1999, p.231 ss.; 1994, p. 383 ss)[3]. Como se advertirá, este planteamiento de la democracia deliberativa se apoya a su vez también en una teoría de la acción social, que fundamenta la prioridad conceptual del uso consenso-comunicativo de la racionalidad, y, como ya se mencionó, en el principio del discurso, que es condición de validez de validez de decisiones y normas adoptadas en el marco de un estado democrático de derecho.

Es precisamente en base a estos fundamentos que cabe analizar el ideal rawlsiano de reciprocidad.

Rawls y el ideal deliberativo de reciprocidad

Si bien ya había sido sugerido en su *Teoría de la justicia* (1971), el ideal deliberativo de reciprocidad fue introducido y desarrollado por Rawls como criterio de justificación de decisiones y normas en un ensayo aparecido en el mismo año con el título de *Justice as reciprocity (*Publicado en Rawls, 1999c, p.190 ss), y más adelante también es tematizado por el autor en su obra de 1993, *Liberalismo político (*RAWLS, 2005, p.17 ss). Este ideal, que desempeña un papel fundamental en el ordenamiento político de una sociedad por su rol en el proceso decisorio, fija los términos justos de cooperación entre individuos libres e iguales mediante ciertas exigencias que estos deben cumplir para la justificación de cuestiones de justicia básica.

Ahora bien, ¿cuál es la operatividad exacta del ideal rawlsiano de reciprocidad, y qué presupuestos comporta? El problema del liberalismo político, señala el autor, "consiste en elaborar una concepción de la justicia política para un régimen constitucional democrático, concepción que la pluralidad de doctrinas razonables pudiera aceptar y suscribir" (RAWLS, 2005, p. 17-18). Ante la pregunta de "¿cómo es posible que pueda existir una sociedad estable y justa de ciudadanos libres e iguales profundamente dividida por

[3] Para un análisis de este tema en la teoría discursiva del derecho de Habermas, véase Habermas, 1994, p. 109 ss.

doctrinas religiosas, filosóficas y morales, aunque razonables, incompatibles entre sí?", o "¿por qué podría una tal concepción política obtener apoyo?" (RAWLS, 2005, p. 17-18), el ideal de reciprocidad de Rawls se orienta a lograr en dicho contexto el asentimiento de los demás interlocutores involucrados recurriendo en parte a un *uso estratégico de la racionalidad*[4]:

> […] los términos justos de cooperación especifican cierta idea de reciprocidad: todos los que participan en la cooperación, y que cumplen con su parte según lo requieran las reglas y los procedimientos fijados, *se beneficiarán de manera apropiada*". [En otros términos, esto significa que] "la reciprocidad es una relación entre ciudadanos expresada mediante principios de justicia que regulan un mundo social en el que *cada cual sale beneficiado.* (RAWLS, 2005, p.17-18, subrayado agregado)[5].

En Rawls los términos justos de cooperación definen su concepción de justicia política, y para lo cual se requiere de una interacción social en donde los ciudadanos, precisamente porque difieren profundamente en sus respectivas doctrinas comprehensivas, intentan lograr apoyo para sus propuestas basándose en un ideal de reciprocidad que (en parte) se centra en la conveniencia mutua: "esta idea se sitúa entre la idea de imparcialidad, que es altruista (pues su motivación es el bien general), y la idea de mutua ventaja, que supone que cada cual tendrá ventajas respecto a su presente o esperada situación futura." (RAWLS, 2005, p.17). Puesto que el autor reconoce el hecho de que a las personas razonables no las motiva el bien general como un fin en sí mismo, sino el deseo de que haya un mundo social en que ellas, en tanto que ciudadanos libres e iguales, puedan cooperar con las demás en términos que todos puedan aceptar, "la reciprocidad debe regir en ese mundo de manera que todos se beneficien." (RAWLS, 2005, p. 49)[6]. Este sentido estratégico del ideal deliberativo de Rawls también está presente en algunos pasajes de su tematización sobre la justicia. En efecto, en su escrito de 1971 sobre "La justicia como reciprocidad." (RAWLS, 1999c, p. 190 ss.) sostiene que la descripción

[4] Como se advertirá más adelante cuando se analice la Teoría de la acción social de Habermas (sección III), el sentido en que se utiliza aquí la expresión "uso *estratégico* de la racionalidad" implica una acción prudencial por parte de un interlocutor, la cual, ante todo, está regida por la consecuente búsqueda de satisfacción de los propios intereses. Se trata pues de una acción que presupone la perspectiva egocéntrica del individualismo, en donde la integración busca *en última instancia siempre* el beneficio individual, aunque las mutuas conveniencias puedan producir formas aparentes de interacción cooperativa (De ZAN, 1993, p.178, ss.).

[5] La tematización rawlsiana de los principios de justicia está en Rawls, 1999a, p.6, 41-42; 2005, p.5; 1999c, p. 193.

[6] No obstante este sentido de ventaja o conveniencia que Rawls menciona aquí, hay que tener en cuenta el hecho de que el autor ubica ante todo el valor de la autonomía y la libertad de los individuos, sin admitir la posibilidad de sacrificar derechos básicos de unos pocos para beneficiar a una mayoría. Esto es importante tenerlo presente aun cuando en ocasiones parece, como en este caso (, y como veremos en otros explícitamente lo reconoce), acercarse a posiciones de tipo utilitaristas a partir de promover un uso estratégico de la racionalidad como guía de las acciones de los sujetos.

hobbesiana de la sociedad a menudo es lo suficientemente cierta para describir las relaciones entre las personas, en especial (pero no exclusivamente) las personas jurídicas, por lo tanto

> se puede formar una concepción más realista de esta sociedad si se piensa en ella como un conjunto de familias, o de alguna otra asociación, [caracterizada por] el mutuo interés propio. Tomando el término 'persona' ampliamente, [...] no es necesario suponer que esas personas estén mutuamente interesadas en todas las circunstancias, pero sin embargo [sí es posible suponer tal cosa] en las situaciones habituales en las que participan en sus prácticas comunes sobre las cuales surge la cuestión de la justicia. (RAWLS, 1999c, p. 198-199).

El ideal deliberativo de reciprocidad implica entonces privilegiar la búsqueda de soluciones a problemas de justicia de modo que estas resulten aceptables para los demás porque pueden beneficiarse. Y el hecho de que tales "buenos argumentos" resulten aceptables para todos, implica excluir aquellas razones que se relacionan con creencias comprehensivas como base para decidir sobre temas de justicia, incluso cuando estas puedan estar correctamente fundamentadas para quienes las sostienen:

> El liberalismo político no es un liberalismo comprehensivo. No adopta una posición general sobre preguntas como '¿acaso el orden moral deriva de una fuente externa (por ejemplo Dios), o surge de la naturaleza humana?', sino que deja que diferentes puntos de vista comprehensivos las contesten a su manera. (RAWLS, 2005, p. 27).

Por esta razón "el hecho de que profesemos determinada doctrina comprehensiva (religiosa, filosófica o moral) no es razón válida para que propongamos, o esperemos que otros acepten, una concepción de la justicia que favorezca a quienes tienen ese credo. [...] Esto sugiere que dejemos a un lado cómo se relacionan estas doctrinas de las personas con el contenido de la concepción política de la justicia". Rawls modela así su concepción de la justicia "colocando las doctrinas comprehensivas de las personas tras el velo de la ignorancia." (RAWLS, 2005, p. 24-25; RAWLS, 1999a, p.11-13, esp. 118 ss).

De acuerdo con esto, algunos autores caracterizan este posicionamiento rawlsiano como asumiendo una "abstinencia epistémica". Garreta Leclerq señala que la pretensión de justificación deliberativa de las políticas fundamentales del Estado a partir de dicho ideal, requiere dejar de lado tales doctrinas religiosas, filosóficas, o morales[7]. Y para A. Gutmann y D. Thompson,

[7] Garreta Leclerq, 2012, p. 291, 290, 292; 2009a, p. 179-184. Por su parte, en su reciente trabajo sobre *El orden de la razón pública*, G. Gaus sostiene la tesis de que los ciudadanos pueden asumir alguna forma básica de razonabilidad sin necesidad de renunciar a las creencias, razonables pero conflictivas, de tipo religiosas o filosóficas que sostienen, pero para esto, afirma, primero sería necesario excluirlas de los procedimientos decisorios sobre temas de justicia (GAUS, 2011, p. 101 ss. -esp. p. 104-117-, p. 170-171).

el ideal de reciprocidad implica que en el plano de la justificación política las diversas posiciones en pugna deben poder realizar afirmaciones en términos que todos puedan aceptar porque todos resultan beneficiados (GUTMAN y THOMPSON, 1996, p. 55 (la cita está en *Garreta Leclerq*, 2009[a], p. 182); se trata, pues, de dar prioridad al desarrollo de propuestas que puedan ser justificadas frente a sus interlocutores. Quien participa de este proceso decisorio, antes de plantear sus pretensiones de validez respecto de la justicia, debe considerar el hecho de que hay argumentos que conviene excluir ya que no serán aceptados como razones válidas para justificar su posición; y por esta razón conviene no tener en cuenta aquellos (argumentos) que deriven de creencias comprehensivas de ningún tipo, ni los aportes que estas puedan realizar para decidir.

Proceder a la fundamentación pública y deliberativa conforme a las exigencias del ideal de reciprocidad presupone, ante todo, un cálculo de tipo estratégico, consistente en evaluar las posibilidades de alcanzar consensos excluyendo ciertas razones como instancias válidas para justificar decisiones; sólo así podrían los interlocutores involucrados lograr acuerdos justos sobre las políticas fundamentales del Estado.

Estas dos cuestiones mutuamente presupuestas que resultan inherentes al ideal rawlsiano de reciprocidad y que hasta aquí se han venido señalando (también asumidas por algunos de sus comentaristas): uso estratégico de la racionalidad, y el tipo de exclusión del proceso decisorio que plantea Rawls respecto de razones comprehensivas, resultan conceptualmente problemáticas si se las analiza desde el marco teórico que determina las condiciones de validez del procedimiento de deliberación racional que establece la teoría de la democracia deliberativa de Habermas para la fundamentación de decisiones y normas, y de la cual forma parte dicho ideal.

El análisis de aquellas dos cuestiones se realiza, respectivamente, en las siguientes dos secciones.

Teoría de la acción social: uso consenso-comunicativo vs. uso estratégico de la racionalidad

La *Teoría de la acción comunicativa* de Habermas (1981) es una teoría de la acción social que centra su interés en el entendimiento comunicativo como procedimiento de coordinación de un tipo especial de interacción que no ha sido adecuadamente analizado por las teorías estándar de la sociología. Con su teoría de la acción comunicativa Habermas se interesa por clarificar el mecanismo en base al cual los actos de habla organizan y regulan las interacciones sociales[8]; se trata a su vez de un análisis que implica un concepto

[8] Para esto el filósofo a su vez se vale del análisis reconstructivo llevado a cabo por la pragmática universal del lenguaje, que explicita las condiciones universales (y por esto inevitables e irrefutables) del entendimiento

amplio de racionalidad que se expresa de manera distinta en cada uno de los diferentes tipos de acciones.

El filósofo distingue entre diferentes orientaciones para la acción y sus correspondientes situaciones a partir de lo que califica como la "versión no oficial de la teoría weberiana de la acción" (HABERMAS, 1995, p. 384 ss.). Así, una acción estratégica, que domina las operaciones económicas del mercado y la economía, es una acción de carácter social orientada al éxito que implica la observancia de reglas de elección racional tendientes a influir en las decisiones de un oponente, y por ello esta tiene lugar en el ámbito de las relaciones interhumanas del mundo social teniendo en cuenta los efectos previsibles de las decisiones propias sobre las decisiones de los otros. Por su parte, una acción es comunicativa, la cual por supuesto también implica una acción social pero orientada al entendimiento, cuando los planes de acción que ella representa no dependen de un cálculo egocéntrico de utilidades, sino de la coordinación de actos de entendimiento. En otros términos, se trata de una relación comunicativa en el mundo social orientada al entendimiento intersubjetivo y a la formación de un consenso racionalmente motivado considerado como válido por los interlocutores involucrados, que hace posible la coordinación no forzada de sus respectivos planes de acción[9]. La distinción clave de este esquema está dada por la diferenciación entre una acción estratégica orientada al éxito, y una acción comunicativa orientada al entendimiento.

Por cierto que no toda acción comunicativa se orienta al *entendimiento*. En el nivel del análisis empírico-descriptivo de la sociología, el tipo de acción estratégica es el tipo de acción dominante en la sociedad moderna, e incluso toda acción social se puede explicar en cierto nivel conforme a este modelo. Pero un análisis filosófico reconstructivo de las condiciones de posibilidad de la interacción humana demuestra que este modelo explicativo es deficiente, y que la acción estratégica y el uso estratégico del lenguaje son derivados y presuponen el uso comunicativo del lenguaje orientado al entendimiento. En efecto, los actos de habla sólo pueden servir al fin de ejercer una determinada influencia sobre el oyente, por ejemplo para satisfacer intereses subjetivos, si ocultan su verdadera intención y se muestran como orientados a lograr un entendimiento racional, no estratégico, con el interlocutor. "Si el oyente no entendiera lo que el

posible. Cabe mencionar que en Habermas estas dos disciplinas (teoría de la acción - pragmática universal) resultan claramente compatibles porque aun cuando pertenezcan a campos epistémicos diferentes, a la vez se encuentran sistemáticamente conectadas. Para un análisis general de las reconstrucciones racionales de la pragmática universal del lenguaje en Habermas (HABERMAS, 1971, p. 23 ss.,137; 1976, p. 307, 310-311, 313; 1983, p. 29-53 (esp. 40); 1995, p. 185-186, 440 ss.; 1990b, p. 94 ss).

[9] Habermas, 1995, p. 384; Cf. De Zan, (1993, p. 173-174). En opinión de McCarthy, la razón fundamental de este planteamiento estriba en que el lenguaje no puede ser comprendido con independencia del acuerdo al que se llega con él, pues el acuerdo es el *telos* inmanente o función del habla (McCARTHY, 1987, p. 333; Habermas, 1995, p. 386-387).

hablante dice [o creyera que quiere engañarlo], este no podría servirse de actos comunicativos para inducirlo a que se comporte de la forma deseada" (HABERMAS, 1995, p. 407). Sobre esta base Habermas distingue entre un acuerdo comunicativamente alcanzado y uno meramente fáctico, entendiendo a aquel en el sentido de un acuerdo que, y a diferencia de este último, no es inducido o determinado por influencias externas, sino que es racionalmente aceptado como válido por los participantes. Hay acuerdos que, de hecho, son acuerdos forzados (este es el ámbito de la *facticidad*), pero esto no cuenta, dice Habermas, como un acuerdo genuino, el cual se basa en convicciones comunes (por lo cual se identifica el sentido de la *validez*); para el autor el empleo del lenguaje orientado al entendimiento es el *original modus* del mismo, respecto del cual el entendimiento indirecto, estratégico, representa un sentido parasitario o derivado, pues ya lo está presuponiendo y haciendo uso de él para sus propios fines de dominación o manipulación de los interlocutores (HABERMAS 1995, p. 384, ss., esp. p. 386-387, 412; 1990b, p. 73, 75, 86, 104; 1999, p. 102-104). Habermas subraya así la importancia del uso comunicativo del lenguaje para la coordinación social como fundamento de este tipo de interacciones.

Precisamente aquí es donde se manifiesta la diferencia fundamental entre el ideal de reciprocidad que sostienen Rawls y sus comentaristas, y el concepto de reciprocidad inherente a la teoría habermasiana de la democracia deliberativa[10]. Se trata de una diferencia (no política sino) filosófica que sitúa a ambos autores en extremos opuestos: de un lado el sentido contractualista de tipo estratégico propio del constructivismo rawlsiano, y del otro el uso consenso-comunicativo de la racionalidad de Habermas, basado en el carácter reconstructivo de su filosofía y expresado en su teoría de la democracia deliberativa. En efecto, mientras que el autor norteamericano (y sus seguidores) prioriza(n) el punto de vista externo a los argumentos presentados, porque al centrarse en las motivaciones de los ciudadanos[11] se limitan las alternativas planteadas según criterios de efectividad de modo que ellas no se relacionen

[10] En realidad en Habermas no cabe hablar de un "ideal", sino de un "principio" de reciprocidad que, como veremos en la sección siguiente (IV), regula la interacción social mediada por discursos prácticos orientados a la resolución argumentativa de pretensiones de validez. El término "idea" no es un término adecuado para dar cuenta de la filosofía habermasiana, toda vez que el mismo refiere a una instancia supraempírica y atemporal que por definición se diferencia radicalmente del "mundo de la vida" (*Lebenswelt*), concepto que por cierto caracteriza tanto a la Filosofía teórica como práctica del autor (HABERMAS, 1995, 1997, 2003).

[11] De hecho ya en su *Teoría de la justicia* Rawls afirma explícitamente que los principios de justicia "son el objeto del *acuerdo* original" por el cual optan "personas libres y racionales *interesadas en promover sus propios intereses*" generales (RAWLS, 1999a, p.6-7). En efecto, en su opinión "la idea principal es que cuando las instituciones más importantes de la sociedad están estructuradas de modo que obtienen el mayor balance neto de satisfacción distribuido entre todos los individuos pertenecientes a ella, entonces la sociedad está correctamente ordenada y es, por lo tanto, justa. *Lo primero que debemos observar es que realmente existe una manera de pensar acerca de la sociedad que hace fácil suponer que la concepción de justicia más racional es la utilitarista*" (RAWLS, 1999a, p. 14-15, subrayado agregado). Una especificación de Rawls (y consecuente aclaración, aunque sin salirse de esta línea) sobre el sentido en que cabe relacionar el utilitarismo con el ideal de reciprocidad está en Rawls, 1999c, p. 218 ss., esp. p. 219.

con doctrinas comprehensivas que puedan impedir la aceptación de tales argumentos, Habermas privilegia la perspectiva interna teniendo en cuenta las razones inherentes a los argumentos presentados, e independientemente de que las mismas se relacionen o no con tales doctrinas: la legitimidad del acuerdo alcanzado depende del desempeño del valor epistémico que estos argumentos (eventualmente) posean, y de la correspondiente "coacción" que ellos puedan ejercer para alcanzar consensos racionalmente motivados en el contexto de un uso consenso-comunicativo de la racionalidad[12].

Rawls aspira a que ningún factor extra-discursivo, como es el caso de las creencias religiosas, influya en la facultad de juzgar de los interlocutores que forman parte del proceso decisorio sobre temas de justicia, y reconoce que "existe la mayor urgencia para que los ciudadanos lleguen a un acuerdo práctico sobre elementos constitucionales esenciales" (RAWLS, 2005, p. 226). Pero precisamente para esto se requiere de un uso de la racionalidad, no estratégico, sino de tipo consenso-comunicativo, actuando de acuerdo a las estrictas condiciones de sentido del discurso argumentativo, que a su vez definen las condiciones de validez de las decisiones así adoptadas. La definición de temas relacionados con la estructura general de gobierno, o con los derechos y las libertades de la ciudadanía[13], por ejemplo, tiene que poder sostenerse sobre acuerdos racionales sólidos que legítimamente le otorguen estabilidad y garanticen su pervivencia a lo largo del tiempo, que es, precisamente, una de las preocupaciones de la teoría política de Rawls.

El principio del discurso y su valor epistémico

El otro aspecto problemático relacionado con el carácter estratégico del ideal rawlsiano de reciprocidad, está dado por la consecuente exclusión de aquellas posturas o creencias (morales, filosóficas, o aun religiosas) que, debido a sus fundamentos comprehensivos, impedirían alcanzar acuerdos en el marco del proceso público de deliberación. El problema aquí es que tal exclusión también se contrapone a los presupuestos teóricos inherentes al procedimiento decisorio que establece la democracia deliberativa, el cual es definido por el principio del discurso argumentativo.

[12] Cf Habermas, (1971, p. 137; 2000, p. 61). También otros autores señalan (en parte críticamente) este carácter estratégico asumido por la teoría de la justicia de Rawls. B. Barry sostiene que la teoría del filósofo norteamericano concibe a la "justicia como ventaja mutua" (BARRY, 1995, p. 31), A. Buchanan refiere a la concepción rawlsiana de la "justicia como auto interés recíproco" (BUCHANAN, 1990, p. 228-30), y para M. Nussbaum no hay "ninguna razón para pensar que Rawls se haya apartado de Hume", en el sentido de quela cooperación debe ser mutuamente ventajosa en comparación con la alternativa del intento de dominación, y que si este no es el caso las partes deben ser excluidas del alcance de la justicia (NUSSBAUM, 2009, p. 61-62).

[13] Precisamente a estas "estructura" o "derechos y libertades" refiere el tipo de "elementos constitucionales" señalado.

Si bien desde un punto de vista comúnmente aceptado el término "discurso" alude a cierto género de oratoria con la cual se espera convencer a un auditorio, el mismo se ha convertido en un término técnico específico de la filosofía contemporánea, sobre todo a partir de Habermas, que entiende "discurso" como un "examen crítico-argumentativo de las pretensiones de validez presupuestas en una afirmación determinada". Tal examen es necesariamente dialógico, y exige ante todo la *simetría* y la correspondiente igualdad de derechos entre quienes participan en él (HABERMAS, 1971, p. 23 ss. (la cita está en MALIANDI, 2006, p. 234).

Al comienzo se afirmó que el principio fundamental de la democracia deliberativa de Habermas no es el principio de la mayoría, sino el principio del discurso[14]. En Habermas (y también en Apel) este principio establece que la legitimidad de las decisiones depende de que ellas puedan ser respaldadas con las correspondientes razones que las fundamentan, y planteadas en el marco de un procedimiento democrático de deliberación llevado a cabo mediante el intercambio y confrontación de argumentos orientados a la obtención de consensos racionalmente motivados, en donde los interlocutores discursivos se guían exclusivamente por la "fuerza" que sólo ejercen los buenos argumentos (HABERMAS, 1995, p. 339-340). En la estructura teórica de la política deliberativael principio del discurso define entonces las condiciones de validez de las decisiones adoptadas en el marco del procedimiento intersubjetivo de deliberación: "válidas son precisamente aquellas normas de acción a las que todos los posibles involucrados puedan asentir como participantes en discursos racionales" (HABERMAS, 1994, p. 138). "Racionales" significa aquí que los interlocutores consideran la corrección de los argumentos analizados en función del contenido de los mismos, por lo cual no es posible excluir a priori ningún argumento, pues esta es, precisamente, una exigencia normativamente fundamentada del principio del discurso[15]. Esto justifica la exigencia de que todo participante en tales procedimientos tenga igual derecho a exponer y hacer valer sus intereses y opiniones, y a que sus razones sean atendidas en base a su fuerza de validez: no cabe advertir a un interlocutor discursivo algo así como "Usted tiene derecho a plantear las pretensiones de validez que crea conveniente, pero para decidir ellas no serán tenidas en cuenta". Como viene sosteniendo desde su *Teoría de la acción comunicativa*, para Habermas "la argumentación tiene por objeto *producir argumentos* pertinentes que convenzan en virtud de sus *propiedades intrínsecas*, [de modo que puedan]

[14] Habermas también sistematiza este principio como principio de universalización de la ética del discurso en otras obras anteriores (HABERMAS, 1983, p. 75-76; 1984, p. 219).

[15] En efecto, leemos en *Facticidad y validez* que "para [...] el principio D es importante que la clase de temas, contribuciones y fundamentos que en cada caso 'cuenten', no sean restringidas a priori" (HABERMAS, 1994, p. 139).

justificar o rechazar las pretensiones de validez" inherentes a ellos (HABERMAS, 1995, p. 48)[16].

Pues bien, de acuerdo con este planteamiento teórico del principio del discurso, dos son las razones (mutuamente presupuestas) por las cuales en la democracia deliberativa no resulta conceptualmente viable el tipo de exclusión que en el marco de su ideal de reciprocidad lleva a cabo Rawls respecto de doctrinas comprehensivas.

La primera señala que si bien tales doctrinas (al menos las religiosas) seguramente impidan alcanzar consensos, a priori tienen el mismo derecho que las demás razones de ser planteadas a lo largo de todo el proceso decisorio. En efecto, definir el marco conceptual de una teoría de la democracia deliberativa, que ciertamente tiene pretensiones normativas, sobre la base de una cuestión fáctica (dificultad de alcanzar consensos) como justificación del rechazo señalado, implica una contradicción con los presupuestos del principio del discurso, según el cual la aceptación o no de las razones planteadas para la solución de un problema depende de su rendimiento *en* el proceso democrático deliberativo, y (nuevamente) por esto ellas se juzgan desde un punto de vista interno (i.e. en función de sus propios méritos) teniendo en cuenta su desempeño para la definición del tema de justicia eventualmente involucrado[17].

Por supuesto, y acorde con el liberalismo político, no se trata de desconocer la necesidad de no basar el diseño de las políticas fundamentales del Estado en doctrinas comprehensivas, sólo sostenidas por algunos ciudadanos, sino de dejar en claro que en una teoría política que justifica la toma de decisiones en base a un procedimiento decisorio de carácter intersubjetivo y deliberativo, tal justificación tiene que resultar del reconocimiento de los presupuestos normativos del principio del discurso señalados, que definen dicho procedimiento, y que mediante discursos prácticos permiten dirimir la

[16] Este también es el argumento que plantea C. Lafont en su último trabajo cuando tematiza respecto de los fundamentos de la democracia deliberativa de Habermas: "According to the deliberative ideal, citizens who participate in public deliberation have the cognitive obligation of *judging the policies* under discussion *strictly on their merits (instead of, say, their self-interest)*. In order to do so they must examine all the relevant reasons and give priority to those reasons that support the better argument, whichever reasons these may turn out to be"(LAFONT, 2013, por aparecer, subrayado agregado). En un trabajo anterior esta filósofa también señaló que tomar las contribuciones religiosas (o cualquier otra) en los asuntos políticos sólo nos obliga a comprometernos seriamente con ellas. Es decir, nos obliga a evaluarlas estrictamente sobre sus méritos (LAFONT, 2007, p. 247, 249, 250).

[17] En este punto hay que señalar que por momentos Rawls parece aceptar una versión, no sólo exclusiva, sino también inclusiva de la razón pública (RAWLS, 2005, p. 225-226). Ahora bien, el problema con este reconocimiento, enteramente correcto desde el punto de vista del concepto de discurso práctico, es que en el planteamiento teórico de Rawls el mismo entra en contradicción con el requisito de neutralidad e igualdad que este filósofo establece en su *Teoría de la justicia* para la posición original, y sobre la cual se sustenta la estructura teórica de su concepción de la justicia (RAWLS, 1999a, p.11-13, 118 ss.; 2005, p. 24-25). Este reconocimiento de Rawls también puede leerse en su artículo "The Idea of Publical Reason Revisited", publicado originalmente en *The University of Chicago Review*, en 1997, y compilado en Rawls 1999b, p. 765 ss. (al respecto véase esp. p. 770-771).

corrección y legitimidad (o no) de tales decisiones. Argumentos con pretensiones normativas pueden (y si corresponde deben) rechazarse en cualquier discusión genuina; pero una tal discusión lo es cuando todo participante tiene el derecho a plantear y justificar sus propuestas frente a los demás, esto significa que sus intereses (o creencias) puedan "ponerse sobre la mesa" intentando demostrar que son legítimos. Esta es la condición que justifica toda posible exclusión de doctrinas comprehensivas del proceso decisorio: solo así, afirma Nino, "puede sostenerse con sentido que el procedimiento intersubjetivo de deliberación racional tiene mayor poder epistémico para ganar acceso a decisiones legítimas que cualquier otro procedimiento de toma de decisiones colectivas" (NINO, 2003, p. 171-172, p. 168).

Esto nos conecta con la segunda razón que justifica el señalamiento acerca de la inviabilidad conceptual del ideal rawlsiano de reciprocidad en el marco de la democracia deliberativa de Habermas. La misma establece que negar la participación en el proceso deliberativo a aquellas doctrinas o creencias cuyos sostenedores pretendan (y crean poder) plantearlas y defenderlas mediante argumentos como medio para contribuir a la justificación de una decisión política, porque ellas sean de carácter moral, filosófico, o religioso, afecta la calidad epistémica de los resultados que puedan obtenerse. En efecto, tal exclusión puede privar al proceso de una perspectiva importante para conocer, por ejemplo, algún aporte relevante para la definición de dicha decisión; como señala Estlund, "el valor epistémico reside en poder inyectar en el proceso la valiosa perspectiva de algunos ciudadanos o individuos reales" (ESTLUND, 2011, p. 295-296, p. 299-300), que en conjunto pueden contribuir a tomar mejores decisiones[18]. De otro modo se priva al acuerdo obtenido de todo momento intelectivo que vaya más allá del cálculo de los propios intereses, y por lo cual el mismo ya no puede ser entendido epistemológicamente en sentido estricto[19].

Estas exigencias teóricas que establece el principio del discurso, a diferencia de las que plantea el ideal de reciprocidad de Rawls, son las que tienen que satisfacerse en una democracia deliberativamente fundamentada, ya que se constituyen en condiciones de validez de las decisiones adoptadas, y también de la justificación de toda posible exclusión de razones. Por esta

[18] Cfr. en este sentido el actual posicionamiento habermasiano sobre la religión, cuando afirma que "no se debe negar a las imágenes de la religión un potencial de verdad", y con el que podrían hacer importantes aportes a las discusiones públicas. En efecto, señala el autor que el Estado "no puede desalentar a los creyentes y a las comunidades religiosas para que se abstengan de manifestarse *como tales* también de una manera política, pues no puede saber si, en caso contrario, la sociedad no se estaría desconectando y privando de importantes reservas para la creación de sentido" (HABERMAS, 2006, p. 138, p. 150; véase también Habermas, 2012, p.238 ss., esp. p. 251-252, 326-327).

[19] Cf. Habermas, (2000: 62, 137). En su debate con Rawls y su concepción del liberalismo político, afirma Habermas que el filósofo norteamericano "quiere asegurar a las afirmaciones normativas –y a la teoría de la justicia en conjunto- un cierto carácter vinculante apoyado en un reconocimiento intersubjetivo fundado, [pero] sin concederle un sentido epistémico" (HABERMAS, 1998, p. 58, 59, 60, 62).

razón no podría afirmarse que la explicitación de los problemas de orden conceptual que en el marco teórico de la democracia deliberativa se plantean con este ideal de reciprocidad de Rawls "comporta un señalamiento poco caritativo"[20]: el criterio fundamental que garantiza la validez y objetividad de los acuerdos, está dado por el reconocimiento de las exigentes condiciones del principio del discurso argumentativo, y no por la obtención de beneficios o ventajas que (si bien dadas ciertas condiciones) Rawls tiene en cuenta con su ideal en cuestión como fundamento para la definición de temas de justicia básica en sociedades plurales con diferentes doctrinas comprehensivas.

Conclusión

El ideal de reciprocidad de Rawls adopta un lugar fundamental en la estructura teórica de su concepción deliberativa de la política. Su propuesta para el diseño y la operatividad general del estado democrático de derecho estriba, ante todo, en la necesidad de apelar a un comportamiento de tipo estratégico por parte de los interlocutores involucrados en el proceso decisorio, por el cual se excluye del mismo a toda doctrina comprensiva que pueda dificultar la posible obtención de acuerdos políticos sobre cuestiones de justicia básica que determinan el diseño institucional. Sin embargo, y desde el punto de vista de la democracia deliberativa de Habermas, este tipo de exclusión que plantea Rawls invierte el orden de prioridad conceptual al privilegiar un uso estratégico de la racionalidad por sobre uno de tipo consenso-comunicativo, cuando en realidad es este último el que permite justificar decisiones sobre temas de justicia, y ello como resultado de la confrontación de argumentos en el marco de un procedimiento intersubjetivo de deliberación racional orientado a la obtención de consensos no forzados; esto permite a su vez fundamentar el valor epistémico de las decisiones así adoptadas en el estado democrático de derecho (aun de aquellas que se plantean para excluir determinas pretensiones de validez debido a la insuficiencia de las razones que las fundamentan). Por esto la exclusión rawlsiana inherente a su ideal de reciprocidad, además de ser conceptualmente inviable, también es fácticamente innecesaria.

Si una teoría de la justicia (como la de Rawls) pretende tener un carácter normativamente vinculante, y contribuir al mejoramiento de la calidad institucional del estado de derecho, entonces tiene que poder asegurar un nivel mínimo de calidad epistémica, que viene dado por el reconocimiento de los presupuestos del principio del discurso argumentativo inherentes al concepto de racionalidad comunicativa, también expresado en la concepción habermasiana de la política deliberativa.

[20] Este es el caso de Lister. Lister, 2011, p. 96.

Es en este marco teórico en donde resulta entonces viable comenzar a analizar la posibilidad de introducir un nuevo argumento en defensa del principio deliberativo de reciprocidad. Este sería el próximo paso.

Referencias bibliografícas

BARRY, Bryan. *Justice as Impartiality*, v. 2, Oxford Political Theory, 1995.

BUCHANAN, Allan. "Justice as Reciprocity versus Subject-Centered Justice", en *Philosophy & Public Affairs*, v. 19, n. 3, 1990, p. 227-252.

COHEN, Joshua. "Deliberation and Democratic Legitimacy". En: BOHMAN, J., REHG, W., *Deliberative Democracy*. Cambridge, Mass., The MIT Press, 1997, p. 67-91.

_____. "Freedom and Money," In: OTSUKA, M. (Ed.). *On the Currency of Egalitarian Justice, and Other Essays*. Oxford: Oxford University Press, 2011.

DE ZAN, Julio. *Libertad, poder y discurso*. Buenos Aires: Almagesto, 1993.

_____. *La ética, los derechos y la justicia*. Montevideo, 2004.

ESTLUND, David. *La autoridad democrática. Los fundamentos de las decisiones políticas legítimas*. Buenos Aires: Siglo XXI, 2011.

G. LECLERQ, Mariano. "Acerca de las dificultades para justificar el ideal deliberativo de reciprocidad". En: GARRETA LECLERQ, Mariano; MONTERO, Julio (Eds.). *Derechos humanos, justicia y democracia en un mundo transnacional*. Buenos Aires: Prometeo, 2009a.

_____. "Democracia deliberativa y justificación mutua". *Revista de Filosofía*, v. 34, n. 2, 2009b, p. 5-27.

_____. "Liberalismo político y reciprocidad: Justificación epistémica de *creencias versus* justificación moral de *acciones*". En: *Isegoría*, n. 46, 2012, p. 279-294.

GAUS, Gerald. *The Order of Public Reason:* A Theory of Freedom and Morality in a Diverse and Bounded World. Nueva York: Cambridge University Press, 2011.

GUTMANN, David; THOMPSON, Amy. *Democracy and Disagreement*, Cambridge. Harvard University Press, 1996.

HABERMAS, Jürgen. *Theorie und Praxis*. Frankfurt: Suhrkamp, 1971.

_____. „Was bedeutet, universalpragmatik'?". En: APEL, K. O. (Ed.). *Sprachpragmatik und Philosophie*. Franfurt: Suhrkamp, 1976.

_____. *Theorie des kommunikativen Handelns (I: Handlungsrationalität und gesellschaftliche Rationalisierung)*. Frankfurt: Suhrkamp, 1995 (por la que se cita), 1981.

_____. *Moralbewusstsein und kommunikatives Handeln*. Franfurt: Suhrkamp, 1983.

_____. "Über Moral und Sittlichkeit - Was macht eine Lebensform 'rational'?", en Schnädelbach, H. (Ed.). *Rationalität*. Frankfurt: Suhrkamp, 1984, p. 218-235.

_____. *Moralbewusstsein und kommunikatives Handeln*. Franfurt: Suhrkamp, 1983.

_____. *Pensamiento postmetafísico*. Madrid: Taurus, 1990b.

_____. *Faktizität und Geltung. Beiträge zur Diskurstheorie des Rechts und des demokratischen Rechtsstaats*. Frankfurt: Suhrkamp,1994.

_____. *Teoría de la acción comunicativa. Complementos y estudios previos*, Madrid: Cátedra, 1997.

_____. "Reconciliación mediante el uso público de la razón", En: HABERMAS, J., RAWLS, J., *Debate sobre el liberalismo político*. Barcelona: Paidós, 1998.

_____. *La inclusión del otro*. Barcelona: Paidós, 1999.

_____. *Aclaraciones a la ética del discurso*. Madrid: Trotta, 2000.

_____. *Acción comunicativa y razón sin trascendencia*. Buenos Aires: Paidós, 2003.

_____. *Entre naturalismo y religión*. Barcelona: Paidós, 2006.

_____. *Nachmetaphysisches Denken II:* Aufsätze und Repliken. Frakfurt: Suhrkamp, 2012.

LAFONT, Cristina. "*Religion in the Public Sphere*: Remarks on Habermas's Conception of Public Deliberation in Postsecular Societies". *Constellations*, v. 14, n. 2, 2007.

_____. "Religious Pluralism in a Deliberative Democracy", Forthcoming In: REQUEJO, Ferran.; UNGUREANU, Camil. (Eds.). *Secular or Post-secular Democracies in Europe? The Challenge of Religious Pluralism in the 21st Century*. London: Routledge, 2013.

LISTER, Andrew. "Justice as Fairness and Reciprocity". *Analyze & Kritik*, v. 33, n. 1, 2011, p. 93-112.

MALIANDI, Ricardo. *Ética:* dilemas y convergencias. Buenos Aires: Paidós, 2006.

MARTINS, Paulo. "La teoría democrática y las bases anti-utilitaristas de la asociación". *Revista argentina de sociología*. Buenos Aires: v. 6, n. 10, jun. 2008, p. 13-33.

MCCARTHY, Thomas. *La teoría crítica de Jürgen Habermas*. Madrid: Tecnos, 1987.

NEUFELD, Blain. "Reciprocity and Liberal Legitimay: Critical Comment on May". *Journal Ethics & Social Philosophy*, Junio 2010. Disponible en: http://www.jesp.org/articles/download/RLL_copyeditedformatted.pdf. Accedido el 1°. abril 2013.

NINO, Carlos Santiago. *La constitución de la democracia deliberativa*. Barcelona: Gedisa, 2003.

NUSSBAUM, Martha. *Frontiers of Justice:* Disability, Nationality, Species Membership. Harvard: Harvard University Press, 2009.

RAWLS, John. *A Theory of Justice*. Harvard: Harvard University Press, 1999a. (por la que se cita).

_____. *Political Liberalism*. Nueva York: Columbia Univesity Press, 2005, (por la que se cita).

_____. *The Law of Peoples with "The Idea of Public Reason Revisited*. Cambridge: Harvard University Press,1999b.

_____. FREEMAN, S. *Collected Papers.* (Ed.). [s.l.]: Harvard University Press, 1999b.

RICO M., Carlos. "Juicio político y virtud cívica en la democracia deliberativa", *Prisma social*, n. 2, junio 2009. Disponible en: <http://www.isdfundacion.org/publicaciones/revista/pdf/n2_4.pdf> Accedido el 1°. abril 2014.

ROBLES, José Manuel. "Cuatro problemas teóricos fundamentales para una democracia deliberativa". *Polis*, v. 7, n. 1, 2011, p. 45-67.

ARGUMENTOS
Revista de Filosofia

RESENHA

José Maurício de Carvalho*

DAU, Shirley e DAU, Sandro. *Ciência*: pesquisa, métodos e normas. Mutum: Expresso, 2013, 173 p.

Os autores são professores de Filosofia e trabalham com Metodologia Científica há anos, havendo publicado diversas obras sobre o assunto. Sandro é Doutor em Filosofia com estágio de pós-doutorado em Ética e Shirley é mestre em Filosofia e estuda Filosofia analítica, Lógica e Linguagem. Dedicam-se também a Epistemologia e História das ciências.

Este livro foi organizado em 16 capítulos e trata da pesquisa científica e sua divulgação. Parte da fundamentação teórica dos estudos científicos para a especificidade das práticas de investigação que contribuem para o desenvolvimento da ciência. Os autores abordam, como atividades práticas, o processo de construção: do projeto e da pesquisa em geral, dos resumos, referências, monografias, relatórios técnico-científicos e artigos científicos.

No capítulo inicial os autores examinam a especificidade dos textos científicos e oferecem técnicas de divisão e interpretação das partes que permitam a boa compreensão desses textos. Esclarecem a diferença entre comentário e explicação, mostrando que a segunda prevalece sobre o primeiro, porque é anterior e restringe-se ao texto, enquanto o comentário pressupõe a explicação, interroga e não se restringe ao texto.

Segue-se o capítulo dedicado aos métodos utilizados na ciência. Eles são apresentados como "caminho estabelecido por determinada ciência, a fim de conseguir conhecimentos válidos por intermédio de instrumentos confiáveis" (p. 17) ou, mais rigorosamente, citando os autores L. Liard, como "conjunto de processos de conhecimento que constituem a forma de uma determinada ciência" (id, p. 17). Neste caso o método possui uma base lógica e dois pontos básicos: reprodutibilidade e falsificabilidade. O primeiro ponto significa que a pesquisa ou experimento pode ser repetido por qualquer pesquisador e o outro que suas hipóteses podem ser recusadas ou falsificadas. As ciências usam variados métodos e, portanto, a escolha do método deve se adequar ao problema e tratamento escolhidos pelo pesquisador. Os autores explicam que a ciência moderna desenvolveu-se partindo da observação, da formulação de hipóteses

* Doutor e professor de Filosofia na UFSJ. Email: *josemauriciodecarvalho@gmail.com*

falsificáveis, da indução ou colocação das hipóteses à prova, da interpretação dos resultados e formulação da teoria, estabelecimento da conclusão.

O capítulo terceiro trata do resultado dos experimentos consolidados em teorias que são fundamentais em qualquer ciência. Os autores aproximam a origem da Ciência com o da Filosofia, remetendo-a ao século VI a. C. na antiga Grécia. Esclarecem que a Ciência moderna, cujo método foi desenvolvido por Galileu Galilei, ganhou perfil diferente e afastou-se da Filosofia e Ciência antes praticadas, adotando "experiências rigorosas, organizadas em leis gerais, por outras palavras, é qualquer corpo de conhecimentos fundado em observações dignas de fé e organizado no sistema de proposições ou leis gerais" (p. 26). Ao tratar da especificidade da Ciência moderna estruturada na razão aplicada (experimental), os autores a diferenciam da Filosofia que se baseia na razão pura; da Religião que usa a fé, possui caráter mais subjetivo e depende da crença de cada um, e da Arte baseada na intuição e não na razão. Eles diferenciam o método indutivo, que vai do particular para o geral, do método dedutivo, que segue o caminho inverso. O conhecimento científico, de modo geral, possui os seguintes traços: "classificar os conhecimentos, descrever os fatos, explicar os fenômenos; interpretar os diferentes casos; ser autocorretivo, experimental, descritivo, particular, cumulativo, operativo" (p. 31). Os autores esclarecem que o conhecimento científico é parcimonioso, isto é, prefere explicações simples. Salientam o propósito da ciência de construir teorias que expliquem os fenômenos e se referem à estas teorias como "uma visão sobre um tema" (p. 40). Eles as caracterizam como possuidoras de "definição rigorosa; coerência interna, generalização por meio de deduções, ampliação do conhecimento" (p. 40). Destacam ainda a utilidade das teorias como critério para sua continuidade.

O capítulo IV é dedicado ao estudo dos conceitos, que são a base da ciência. Contudo, comentam os autores, que não basta reunir conceitos para chegar a uma ciência, é preciso que os conceitos estejam organizados em teorias, cujas características foram examinadas no capítulo anterior. O conhecimento adequado dos conceitos é um bom critério para saber se se conhece uma teoria e para organizar o debate, pois "aquele que afirma e aquele que pergunta, devem ter claro qual o significado do conceito empregado" (p. 44). Um conceito científico caracteriza-se "por ter um significado claro, preciso e abstrato, que não resulta de preferências, de gostos e de anseios individuais" (p. 44). Contudo, a característica principal de um conceito científico é "identificação dos elementos centrais daquilo que é estudado" (p. 45) e não se referir "a um caso único, a um fenômeno, mas a classes, a grupos, a relações, etc." (p. 45). A compreensão de um grupo de conceitos que estão numa ciência dirige o olhar do cientista, assim ele perceberá o fenômeno com os olhos da ciência que lhe é familiar, ou "cada um perceberá o mundo com os conceitos que for condicionado" (p. 46).

Segue-se um capítulo dedicado a verdade. Os autores tratam dos critérios de verdade estabelecidos por três grandes escolas filosóficas: realismo (relação entre a consciência e a coisa), idealismo (coerência interna) e pragmatismo (utilidade das afirmações). No que se refere propriamente às verdades científicas, elas se dividem em três tipos: as que se limitam a formular o que pode ser exato, as que aceitam a probabilidade e aquelas que se abrem à possibilidade do mundo de cada pessoa influir no conhecimento. Estas últimas são as ciências humanas que, com base na fenomenologia, reconhecem, por baixo dos fatos objetivamente descritos, elementos do chamado mundo da vida. No que se refere à relação dos fatos, os autores distinguem as verdades lógico-formais, das objetivas, ontológicas e morais. Poderíamos entender as primeiras como as da Lógica e Matemática, as segundas como as das Ciências da natureza e as duas últimas da Filosofia. Quanto a relação com a verdade ela pode produzir quatro tipos de dúvidas: "espontânea, refletida, metódica e universal." (p. 53). A primeira é aquela em que a pessoa não emite juízo sobre algo, mesmo que tenha elementos para fazê-lo, a segunda nasce da ausência de elementos necessários à conclusão, a terceira é um tipo de método empregado para chegar a uma verdade indubitável e a última refere-se à posição dos céticos, que negam a possibilidade de chegar a verdades fundamentais. Quanto aos critérios de verdade os autores apontam seis: a autoridade (abandonada pelas ciências modernas), a evidência (o que aparece para o indivíduo), o senso comum (uma espécie de instinto comum), a necessidade lógica (ausência de contradição) e a experiência.

O capítulo VI estuda a pesquisa científica e os autores a definem como a que utiliza "métodos racionais, científicos, na comprovação ou não, das teorias apresentadas" (p. 57). Eles apresentam como objetivo da pesquisa científica "encontrar respostas coerentes, para os problemas (questões) propostos pelo pesquisador" (p. 58) e diferenciam as pesquisas quantitativas, que trabalham com a quantificação das variáveis, das qualitativas, mais empregadas nas ciências humanas. Esclarecem que essas últimas se desenvolveram no século passado e resumem o debate então realizado entre os intérpretes das chamadas ciências duras e as outras. Tratam, ainda que superficialmente, dos limites da razão experimental que foi desenvolvida na modernidade para tratar dos problemas do homem, assunto da fenomenologia.

A partir do capítulo VII o livro ganha um caráter prático, orientando o leitor em como utilizar as técnicas empregadas na pesquisa científica. Distinguem esquema, resumo e fichamento. Esclarecem que esquema "permite ao estudante compreender uma obra em seu todo" (p.63). Eles propõem no capítulo VIII o resumo como "apresentação concisa de um texto qualquer" (p. 65) e tratam no capítulo IX de um tipo especial de Resumo denominado Resenha. Explicam que Resenhas são um tipo específico de resumo seguido de comentário crítico, por isto dizem que eles devem ser "elaborados por especialistas"

(p. 67). Normalmente as revistas científicas recebem bem este tipo de resumo dedicando-lhe uma parte própria, porque é importante aos especialistas da área terem um resumo comentado das obras daquela ciência pois não é possível, hoje em dia, ler tudo que se publica nas diversas áreas da ciência.

O capítulo X é dedicado ao fichamento, definido como técnica para "guardar um grande número de informações sobre um documento em pequeno espaço." (p. 69). O fichamento pode ser da obra toda ou de uma parte, além de conter citações que serão úteis na elaboração do trabalho que pretende fazer. O capítulo XI é um resumo da NBR 10520 e explica como fazer citações curtas e longas. Segue-se um capítulo sobre como fazer referências, resumo da NBR 6023, de livros, monografias, periódicos, eventos, trabalhos em eventos, legislação, jurisprudência, doutrinas, filmes, documentos cartográficos e sonoros.

O capítulo XIII explica como se faz um projeto de pesquisa, apresentado como "caminho que será percorrido, no estudo do problema proposto" (p. 97). Os autores detalham os elementos imprescindíveis do projeto (capa, folha de rosto, sumário, apresentação, justificativa, área de concentração, natureza, delimitação do assunto, revisão da literatura, problema, hipóteses, procedimento, análise dos dados, objetivos, conteúdo, metodologia, cronograma, referências, anuência do orientador).

O capítulo seguinte é dedicado à monografia, definida como "texto sobre um único assunto" (p. 129), e que pode ser desde um TCC até uma tese de doutoramento. Consiste num resumo da NBR 14724. Suas características básicas são: sistematicidade, metodologicidade e relevância. Uma monografia se divide, geralmente, em cinco partes: "introdução, desenvolvimento, conclusão, bibliografia, notas." (p. 131). Os autores finalmente afirmam que a estrutura formal da monografia são três grandes partes: os elementos pré-textuais, os textuais e os pós-textuais.

Os capítulos finais explicam como fazer um relatório e um artigo científico, respectivamente comentando as NBRs 10719 e a 6022. Os artigos científicos, matéria do último capítulo, são definidos como: "publicação com autoria declarada, que apresenta e discute ideias, métodos, técnicas, processos e resultados nas diversas áreas do conhecimento." (p. 163). O artigo serve para divulgar um tema estudado e deve vir em linguagem "clara, coerente, objetiva, impessoal" (p. 163) e conter os elementos pré-textuais, textuais, pós-textuais.

Este livro é importante porque coloca o leitor diante do fato de que fazer ciência é mais do que aprender conceitos e teorias, exige produzi-la. Esta atitude é própria de um tipo de ciência desenvolvido na modernidade, com os estudos de Francis Bacon, Isaac Newton e os iluministas, que contrapunham a nova formulação da ciência à antiga construída na velha Grécia por Platão e Aristóteles. A ciência moderna nunca está pronta, mas em continuado processo de construção. Por isso, o estudo das técnicas de pesquisa é essencial

numa atividade que está sempre se fazendo. O movimento iluminista reforçou a confiança na razão aplicada e no modelo de ciência moderna pautada na observação dos fatos, experiência e cálculos. Os autores incorporam aspectos importantes dos estudos de filosofia da ciência. Entre eles o entendimento de que a enumeração dos fatos ou descrição dos conceitos não é suficiente para fazer ciência, antes é preciso comparar os fatos observados e julgá-los para construir teorias. Isto é o que ensinava, por exemplo, o médico e fisiologista francês Claude Bernard no século XIX. Os autores incorporaram ainda uma concepção mais atual de ciência que trata da sua validade em virtude da autocorreção, princípio baseado na falibilidade das teorias, conforme postulado por Charles Sanders Peirce e pela falsificabilidade, conceito desenvolvido por Karl Popper. Para este último uma teoria é válida não devido a sua demonstração, mas por sua permanência provisória, enquanto não vingam os esforços por refutá-la.

No capítulo IX os autores tocaram numa questão importante da ciência moderna, a sua necessária especialização. Como lembra Ortega y Gasset no capítulo XII de *La rebelión de las masas* (O.C., Madrid, Alianza, v. IV, 1994): "nem sequer a ciência empírica, tomada em sua integridade, é verdadeira se separada da Matemática, da Lógica, da Filosofia. Porém o trabalho em que nela se tem, irremediavelmente, tem que ser especializado." (p. 217). Esta especialização exigida pela ciência moderna contém, contudo, um grave risco que o Ortega repetiu em mais de um lugar e isto não foi mencionado no livro. É que a especialização não legitima o conhecedor de uma ciência opinar sobre outros assuntos. Quando ele assim faz torna-se uma espécie de novo bárbaro, detalhadamente estudado por Ortega. Este especialista deverá passar por uma reciclagem, se estiver correto o que diz Ortega na continuidade do livro, pois se:

> o especialismo tornou possível o progresso da ciência experimental durante um século, aproxima-se uma etapa nova em que ele não poderá avançar por si mesmo se não encarregar uma geração melhor de construir um novo aparelho mais poderoso. (p. 219-220).

Este novo especialista é uma exigência dos nossos dias, mas ainda assim será ele um especialista.

O assunto nuclear, da perspectiva epistemológica, consiste na discussão entorno à verdade levada a cabo no capítulo V. E aí também há virtudes. Parece importante a distinção dos conceitos de verdade construídos por diferentes escolas filosóficas: realismo, idealismo e pragmatismo. Também parece fundamental a diferença entre os critérios de verdade adotados por diferentes ciências: a verdade exata da linguagem matemática, as afirmações aproximativas da estatística e as verdades cuja objetividade relativa está em disputa com as referências subjetivas do mundo da vida. Faltou indicar que essas dife-

rentes visões de verdade científicas nascem em diferentes tipos de ciência, as primeiras da natureza e as últimas do homem. A distinção entre os diferentes tipos de dúvida também foi muito criativo. O que ficou a merecer maior aprofundamento é o fato de que as verdades são diferentes nas Ciências, na Filosofia, na Religião e até as Pessoais. Todas as formas de verdade são importantes, mas se organizam em níveis distintos. Neste aprofundamento sobre as diferentes verdades faltou também um esclarecimento sobre os limites das chamadas ciências duras, ou a ciência experimental, pois suas teorias começaram a ser refutadas pelo desenvolvimento da própria filosofia da natureza no século que passou.

ARGUMENTOS
Revista de Filosofia

Marcos Silva*

Notes on Thompson's "Wittgenstein on phenomenology and experience". University of Bergen Press, 2008.

I read Thompson's well-written and relevant book 'Wittgenstein on Phenomenology and Experience', published by the University of Bergen Press in 2008, with great interest. My PhD Dissertation, defended in 2012, has direct connections with his main object of investigation, especially because one of my interests there was to evaluate logical problems with the expressiveness of color exclusion within the tractarian background.

Thompson's treatment of the so-called Middle Wittgenstein period, documented by the transitional material that appeared in the *Nachlass,* is for this reader the most seminal feature of his work on Wittgenstein's phenomenology. His commentary provides a useful addition to the leading and influential researchers already focusing on this challenging and oft-neglected material. Thompson manages to handle significant problems with Wittgenstein's exposition about experience and phenomenology without lapsing into the sort of misleading labels and programmatic vagueness that has dominated commentaries of the last two decades in the "Wittgensteinian scholarship", for instance discussions of the tractarian passage 6.53, which orientates the contention of resolute reading. The secondary literature has too often rendered Wittgenstein an isolated and aptly neglected author in contemporary analytic philosophy.

One potentially misleading feature of Thompson's exposition, however, is the symmetric approach that he takes towards presenting Wittgenstein's thoughts about experience and phenomenology; on the contrary, a careful reading seems to reveal that phenomenology was a centrally important topic in Wittgenstein's philosophical development, while experience was not. Consider the frequency and centrality with which phenomenology was directly discussed by Wittgenstein, while any discussion of experience was very often fragmentary and marginal. Moreover, note the kind of association which Thompson draws between the mystical experience in *Tractatus Logico-Philosophicus* [hereafter TLP] as a trigger for the rise of phenomenology in the

* Pós-doutorando em Filosofia na Universidade Federal do Ceará (UFC)/CAPES-PNPD. E-mail: marcossilvarj@gmail.com

transitional period. If Thompson is correct, then the relation is by no means obvious and straightforward, and it deserves a fuller explication. I do agree that some germs of the phenomenology found in Wittgenstein's Middle Period can be already seen in the *Tractatus*, but not in its contention on mystical experience, as Thompson defends, but already in the very beginning of his first book.

Arguably, Thompson's work overlooks the importance of colors and their logical organization in this transitional material. In some passages of *Philosophische Bemerkungen* [hereafter] PB, for instance §81-83, and in some entries of the discussions presented in *Wittgenstein und der Wiener* Kreis [hereafter WWK], such as 'Farbsystem' and 'Die Welt ist rot', Wittgenstein does draw attention to his uses of colors in TLP directly connected to his new phenomenology. I am not talking about the obvious problem in 6.3751, first pointed out by Ramsey (who was not mentioned in any part of Thompson's book). Criticizing this *Tractarian* passage, Ramsey (1923) discovered the *Sackgasse* for the tractarian logic: Some necessary consequences are not due to tautologies. However, I prefer to read this contention through its dual: Some (logical) exclusions are not due to contradictions (but due to contrarieties). My point is that if we read carefully the first two mentions of colors in Wittgenstein's *Tractatus*, namely 2.0131 and 2.0252, which both occur in the work's so-called ontological section, we will see that already some phenomenology was to be expected even there. The italics in 2.0131 strongly suggest a kind of exclusion, surprisingly underdeveloped by Wittgenstein at that time. As this passage 2.0131 already suggests, these italics are not just to be found in color system. The 'etcetera' in this very same passage suggests the multiplicity of 'logical spaces' or 'Satzsysteme', whose treatment are ubiquitous in his "phenomenological" period and given a full treatment.

Another concern might be raised about Thompson's neglect of Ramsey's relevance to Wittgenstein's abandonment of the thesis of the independence of elementary propositions/*Sachverhalt*. Many authors have been said to have influenced Wittgenstein directly or indirectly throughout his carrier. But none of them made a complicated trip from England to Austria, more specifically, to a small village in *Niederösterreich* in the middle of nowhere, to meet personally with Wittgenstein to discuss some (obscure) problems in his (obscure) book. Ramsey was the first one to recognize the significant problem of logical organization that colors posed and the challenge they represented for the *tractarian* logic and image of language. Moreover, as an illustration of a very interesting case of historical completeness, Ramsey already pointed out the color problem within the tractarian philosophy in 1923; he therefore probably anticipated, in 1927, Wittgenstein's later solution for the problem introducing additional rules, pragmatism and games, by using a metaphor of chess. And Ramsey had proposed all of that three years before Wittgenstein had begun

talking significantly about games! The importance of recognizing Ramsey's criticism and his impact on Wittgenstein's solutions in the *Tractatus* is not just a matter of scholarly integrity; it is also a matter of illuminating accurately the conceptual development of key contributions made to logic and mathematics which have become associated with early analytic philosophy.

The total neglect of WWK in Thompson's book, which purportedly intends to unveil Wittgenstein key shifts, is also hard to comprehend. WWK was neither written nor edited by Wittgenstein; yet it is a great historical and philosophical document for understanding the kinds of problem Wittgenstein was dealing with and reacting to in his philosophical development. If the problem is that WWK is not well edited, that can always be established by a careful comparison with Wittgenstein's *Nachlass*. Such an exercise would like reveal that many arguments, metaphors and concepts are indeed very similar. Thompson ought to justify why he very often used PB and not WWK at all. Moreover, in WWK we can see diachronically how things evolved, while, with PB, Rush Ree's interventions make this kind of genetic investigation impossible.

Perhaps also as consequence of not using WWK, Thompson seems to have overlooked the importance of the year 1930 for Wittgenstein's treatment of phenomenological problems. For instance, in the beginning of 1930, the notion of normativity, which is not explored in Thompson's book, arose in Wittgenstein's discussions with Waismann about the number π and the role of axioms in geometry. Another example is the role of June of 1930. At this time, Wittgenstein was preparing Waismann to represent him in a brilliant round table on the nature of mathematics in Könisberg, in which Von Neumann, Carnap and Heyting would participate. In the entry 'Was wäre es zu sagen in Könsisberg' in WWK, we can see both Wittgenstein and Waisman discussing *Grundgesetze*'s criticism of formalism. This entry shows that Wittgenstein defended clearly, against Frege, that formalists are right in holding mathematics as a game. This discussions on formalism also marks Wittgenstein`s decreasing interest in his short-lived phenomenology. In this way, this entry should have played a relevant role in Thompson's evaluation of Wittgenstein's phenomenology.

Another conspicuously absent omission in Thompson's book was some detailed discussions of verificationism. Maybe this is also due to his neglect of WWK in his critique; for it is there that this topic is raised at several points in conjunction with phenomenology. These discussions are important to understand Wittgenstein's influence on Carnap and the Vienna Circle; moreover, the prominence of Wittgenstein's treatment of these two conjoined topics is critical to appreciating the influence on Wittgenstein of Brouwer's intuitionism and revisionism about the role and nature of logic. Thompson does mention, but does not explore in much detail, the clear connection between verificationism and problems with the restrictiveness of truth-functionality. In some way this discussion may link with the reasons why the kind of realist truth theory

defended in the *Tractatus* (based on the notion of sense as truth conditions) must be abandoned. It might be argued that this consequence is directly linked to the full ascendancy of Wittgenstein's phenomenology: 'sense' resolves finally into the concern for finding a method for verification, and not a matter of concern for determining logical truth conditions. Thus a very important key to the role that his phenomenology played in Wittgenstein's official return to philosophy has been neglected, in an otherwise compelling overview of his phenomenology and the notion of experience.

Thompson made, in spite of these problems pointed above, some brilliant remarks on the failure of using calculus to understand human language discussing its lack of determinedness, rigidity (i.e. the well structuredness of rules) and completeness. I recommend Thompson's book to people interested in an introduction to Wittgenstein's (short-lived) phenomenology and for anyone who will profit from sharp, effective criticism of the limitations of the so-called resolute reading.

References

FREGE, Gottlob. *Grundgesetze der Arithmetik.* Band: II. Jena: Verlag Hermann Pohle, 1903.

RAMSEY, Frank. Critical Notes *to Tractatus* Logico-Philosophicus. *Mind*, p. 465-478 ,1923

_____. Facts and propositions. Proceedings of the Aristotelian Society, Supplementary Volumes, v. 7, *Mind*, Objectivity and Fact, p. 153-206, 1927.

SILVA, Marcos. Muss Logik für sich selber sorgen? On the Color Exclusion Problem, the truth table as a notation, the Bildkonzeption and the Neutrality of Logic in the Collapse and Abandomnent of the Tractatus. PHD Thesis - Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, 2012.

WITTGENSTEIN, Ludwig. *Philosophische Bemerkungen*. Werkausgabe Band 2. Frankfurt am Main: Suhrkamp, 1984.

_____. *Some Remarks on Logical Form*. Proceedings of the Aristotelian Society, Supplementary Volumes, v. 9, Knowledge, Experience and Realism p. 162-171 Published by: Blackwell Publishing on behalf of The Aristotelian Society, 1929.

_____. *Tractatus* Logico-philosophicus. Tagebücher 1914-16. Philosophische Untersuchungen. Werkausgabe Band 1. Frankfurt am Main: Suhrkamp, 1984.

_____. Wittgenstein und der Wiener Kreis. Werkausgabe Band 3. Frankfurt am Main: Suhrkamp, 1984.