



Editora
UFPel



Novos Ensaios de Filosofia Analítica

Ensaios Vencedores do Prêmio da Sociedade
Brasileira de Filosofia Analítica

Waldomiro J. Silva Filho
(Organizador)

DISSERTATIO
FILOSOFIA

**NOVOS ENSAIOS DE FILOSOFIA ANALÍTICA: ENSAIOS
VENCEDORES DO PRÊMIO DA SOCIEDADE BRASILEIRA DE
FILOSOFIA ANALÍTICA**

Série Dissertatio Filosofia

**NOVOS ENSAIOS DE FILOSOFIA ANALÍTICA: ENSAIOS
VENCEDORES DO PRÊMIO DA SOCIEDADE BRASILEIRA DE
FILOSOFIA ANALÍTICA**

Waldomiro J. Silva Filho
(Organizador)

DISSERTATIO
FILOSOFIA
Pelotas, 2018

REITORIA

Reitor: Pedro Rodrigues Curi Hallal

Vice-Reitor: Luís Isaias Centeno do Amaral

Chefe de Gabinete: Taís Ullrich Fonseca

Pró-Reitor de Graduação: Maria de Fátima Cossio

Pró-Reitor de Pesquisa e Pós-Graduação: Flávio Fernando Demarco

Pró-Reitor de Extensão e Cultura: Francisca Ferreira Michelin

Pró-Reitor de Planejamento e Desenvolvimento: Otávio Martins Peres

Pró-Reitor Administrativo: Ricardo Hartlebem Peter

Pró-Reitor de Infraestrutura: Julio Carlos Balzano de Mattos

Pró-Reitor de Assuntos Estudantis: Mário Renato de Azevedo Jr.

Pró-Reitor de Gestão Pessoas: Sérgio Batista Christino

CONSELHO EDITORIAL DA EDITORA DA UFPEL

Presidente do Conselho Editorial: João Luis Pereira Ourique

Representantes das Ciências Agrônomicas: Guilherme Albuquerque de Oliveira Cavalcanti

Representantes da Área das Ciências Exatas e da Terra: Adelar José Strieder

Representantes da Área das Ciências Biológicas: Marla Piumbini Rocha

Representante da Área das Engenharias e Computação: Darci Alberto Gatto

Representantes da Área das Ciências da Saúde: Claiton Leoneti Lencina

Representante da Área das Ciências Sociais Aplicadas: Célia Helena Castro Gonsales

Representante da Área das Ciências Humanas: Charles Pereira Pennaforte

Representantes da Área das Linguagens e Artes: Josias Pereira da Silva

EDITORA DA UFPEL

Chefia: João Luis Pereira Ourique (Editor-chefe)

Seção de Pré-produção: Isabel Cochrane (Administrativo)

Seção de Produção: Gustavo Andrade (Administrativo)

Anelise Heidrich (Revisão)

Ingrid Fabiola Gonçalves (Diagramação)

Seção de Pós-produção: Madelon Schimmelpfennig Lopes (Administrativo)

Morgana Riva (Assessoria)

CONSELHO EDITORIAL

Prof. Dr. João Hobuss (Editor-Chefe)
Prof. Dr. Juliano Santos do Carmo (Editor-Chefe)
Prof. Dr. Alexandre Meyer Luz (UFSC)
Prof. Dr. Rogério Saucedo (UFSM)
Prof. Dr. Renato Duarte Fonseca (UFSM)
Prof. Dr. Arturo Fatturi (UFFS)
Prof. Dr. Jonadas Techio (UFRGS)
Profa. Dra. Sofia Albornoze Stein (UNISINOS)
Prof. Dr. Alfredo Santiago Culleton (UNISINOS)
Prof. Dr. Roberto Hofmeister Pich (PUCRS)
Prof. Dr. Manoel Vasconcellos (UFPEL)
Prof. Dr. Marco Antônio Caron Ruffino (UNICAMP)
Prof. Dr. Evandro Barbosa (UFPEL)
Prof. Dr. Ramón del Castillo (UNED/Espanha)
Prof. Dr. Ricardo Navia (UDELAR/Uruguai)
Profa. Dra. Mónica Herrera Noguera (UDELAR/Uruguai)
Profa. Dra. Mirian Donat (UEL)
Prof. Dr. Giuseppe Lorini (UNICA/Itália)
Prof. Dr. Massimo Dell'Utri (UNISA/Itália)

COMISSÃO TÉCNICA (EDITORAÇÃO)

Prof. Dr. Lucas Duarte Silva (Diagramador)
Profa. Luana Francine Nyland (Assessoria)

DIREÇÃO DO IFISP

Prof. Dr. João Hobuss

CHEFE DO DEPARTAMENTO DE FILOSOFIA

Prof. Dr. Juliano Santos do Carmo

Série Dissertatio Filosofia

A Série Dissertatio Filosofia, uma iniciativa do Núcleo de Ensino e Pesquisa em Filosofia (sob o selo editorial NEPFIL online) em parceria com a Editora da Universidade Federal de Pelotas, tem por objetivo precípua a publicação de estudos filosóficos relevantes que possam contribuir para o desenvolvimento da Filosofia no Brasil nas mais diversas áreas de investigação. Todo o acervo é disponibilizado para download gratuitamente. Conheça alguns de nossos mais recentes lançamentos.

Estudos Sobre Tomás de Aquino

Luis Alberto De Boni

Do Romantismo a Nietzsche: Rupturas e Transformações na Filosofia do Século XIX

Clademir Luís Araldi

Didática e o Ensino de Filosofia

Tatielle Souza da Silva

Michel Foucault: As Palavras e as Coisas

Kelin Valeirão e Sônia Schio (Orgs.)

Sobre Normatividade e Racionalidade Prática

Juliano do Carmo e João Hobuss (Orgs.)

A Companion to Naturalism

Juliano do Carmo (Organizador)

Ciência Empírica e Justificação

Rejane Xavier

A Filosofia Política na Idade Média

Sérgio Ricardo Strefling

Pensamento e Objeto: A Conexão entre Linguagem e Realidade

Breno Hax

Agência, Deliberação e Motivação

Evandro Barbosa e João Hobuss (Organizadores)

Acesse o acervo completo em:

wp.ufpel.edu.br/nepfil

© **Série Dissertatio de Filosofia, 2018**

Universidade Federal de Pelotas
Departamento de Filosofia
Núcleo de Ensino e Pesquisa em Filosofia
Editora da Universidade Federal de Pelotas

NEPFil online

Rua Alberto Rosa, 154 – CEP 96010-770 – Pelotas/RS

Os direitos autorais dos colaboradores estão de acordo com a Política Editorial do NEPFil online. As revisões ortográfica e gramatical foram realizadas pelos autores e tradutores.

Primeira publicação em 2018 por NEPFil online e Editora da UFPel.

Dados Internacionais de Catalogação

N170 Novos Ensaios de Filosofia Analítica: Ensaios Vencedores do Prêmio da Sociedade Brasileira de Filosofia Analítica.
[recurso eletrônico] Organizador: Waldomiro J. Silva Filho – Pelotas:
NEPFIL Online, 2018.
355p. - (Série Dissertatio Filosofia).
Modo de acesso: Internet
<wp.ufpel.edu.br/nepfil>
ISBN: 978-85-67332-62-8

1. Filosofia Analítica. 2. Prêmio SBFA I. Filho, Waldomiro J. Silva.

COD 170



Para maiores informações, por favor visite nosso site wp.ufpel.edu.br/nepfil

APRESENTAÇÃO: O MOVIMENTO DA FILOSOFIA ANALÍTICA NO BRASIL HOJE

Este livro reúne quatro capítulos escritos por jovens filósofos brasileiros. Eles foram os vencedores da primeira edição do Prêmio da Sociedade Brasileira de Filosofia Analítica (SBFA), anunciado durante a V Conferência da SBFA, realizada em Salvador, Bahia, entre 10 e 14 de setembro de 2018.

Concorreram ao prêmio doutorandos e jovens filósofos em início de carreira que submeteram artigos e ensaios inéditos que poderiam compreender qualquer uma das áreas da Filosofia Analítica – Epistemologia, Lógica, Filosofia da Linguagem, Ética etc. Para compor a comissão julgadora, a direção da SBFA convidou filósofos reconhecidos pela comunidade filosófica brasileira: Adriano Naves de Brito (UNISINOS), André Leclerc (UnB), Eleonora Orlando (Universidad de Buenos Aires, SADAF), Guido Imaguire (UFRJ), Luiz Carlos Pereira (PUC-Rio, UERJ), Marco Ruffino (UNICAMP), Oswaldo Chateaubriand (PUC-Rio) e Sofia Stein (UNISINOS).

Além dos quatro vencedores, este volume inclui o capítulo “Porchat, neopirronismo e a filosofia analítica” escrito por Plínio Junqueira Smith. Osvaldo Porchat Pereira, falecido em 2017, foi o homenageado da “V Conferência da SBFA” e esse capítulo é uma versão da palestra de abertura do evento. Porchat desempenhou dois papéis fundamentais para a filosofia analítica brasileira. De um lado, fundou o Centro de Lógica e Epistemologia (CLE) na UNICAMP, criando as bases institucionais para o seu florescimento. De outro, dialogou de maneira rica e fértil com vários dos principais filósofos analíticos ao longo de sua carreira. A princípio, George E. Moore e Peter F. Strawson foram referências centrais para a elaboração da sua filosofia da visão comum do mundo. Depois, com o advento do seu neopirronismo, filósofos como Willard v. Quine e Ludwig Wittgenstein se tornaram os mais importantes. Em particular, Porchat desenvolveu uma doutrina neopirrônica da verdade, recorrendo a John L. Austin, e teceu reflexões bastante originais sobre a crítica contemporânea à legitimidade dos problemas filosóficos. Segundo Plínio Smith, Porchat foi se tornando cada vez mais um filósofo analítico, dedicando-se sobretudo aos grandes nomes. Um aspecto importante da apropriação da filosofia analítica, foi o abandono de seu estruturalismo inicial e a sua ênfase em que deveríamos fazer história da filosofia em benefício do filosofar.

Na obra de Porchat, portanto, encontramos duas filosofias que, cada uma à sua maneira, aclimatam discussões clássicas dos filósofos analíticos ao nosso contexto e à nossa língua.

O segundo capítulo, “Strong externalism about reasons” de Rafael G. Vogelmann trata do importante debate internismo/externismo acerca de razões. O autor propõe uma nova perspectiva sobre o internismo (*State Internalism*), tese segundo a qual se um agente tem uma razão para agir de um certo modo, esse agente está comprometido a agir assim sobre a pena de ser declarado irracional. Essa tese tem sido, de um lado, considerada trivial e, do outro, considera-se que ela leva a consequências indesejadas. Para muitos autores, combinado com outras teses plausíveis relativas à racionalidade (a saber, que a racionalidade é puramente processual e que as atitudes correspondentes às demandas morais não são processualmente acessíveis a todos os agentes), o internismo representa uma ameaça à ideia de que todo agente tem uma razão para agir quando ele é moralmente obrigado a agir. O objetivo deste capítulo é argumentar que o internismo não é uma trivialidade e que podemos evitar essa ameaça rejeitando-a. Ao fazer isso, Vogelmann se compromete com o externismo forte (*Strong Externalism*), de acordo com o qual um agente pode falhar em responder a suas razões sem qualquer falha na racionalidade. Seu argumento defende a questão de se todos temos motivos para nos comportarmos moralmente não pode ser resolvida por considerações sobre racionalidade, nem mesmo se aceitarmos a ideia de que a racionalidade é puramente processual e que, em alguns casos, não basta um raciocínio processual correto para que um agente desenvolva uma atitude moral.

O terceiro capítulo, “Color terms and perception” de Raquel Albieri Krempel trata da relação entre termos referentes a cor e percepção. A autora salienta que há muito se sabe que as línguas naturais diferem na maneira como categorizam as cores e, por isso, é razoável questionar se a língua que alguém fala pode influenciar o modo como ela percebe as cores. E, supondo que sim, uma outra questão é como exatamente a língua influencia a percepção de cores. Na primeira parte do capítulo, Krempel considera aspectos históricos do problema de determinar como os termos de cor se relacionam com a percepção de cores. Em seguida, ela apresenta pesquisas empíricas recentes que demonstram que as línguas naturais afetam diretamente a forma como alguém discrimina as cores. Na terceira seção do capítulo, a autora argumenta que essas descobertas podem

receber uma interpretação diferente, a saber, que os conceitos de cores e não as palavras referentes a cores interferem no desempenho da pessoa nas tarefas de discriminação de cores. Finalmente, na seção final, são discutidas as implicações que essas descobertas têm para o tópico filosófico tradicional da relação entre percepção e cognição.

Anderson Luis Nakano é o autor do quarto capítulo, “Wittgenstein, formalism and symbolic mathematics”. Nakano parte de um ensaio recente de Sören Stenlund (“On the Origin of Symbolic Mathematics and Its Significance for Wittgenstein's Thought”) que tenta alinhar a abordagem de Wittgenstein às fundações e à natureza da matemática com a tradição da matemática simbólica. A caracterização da matemática simbólica feita por Stenlund, segundo a qual a matemática é logicamente separada de suas aplicações externas, aproxima-a da posição formalista – e isso levanta naturalmente a questão de saber se Wittgenstein assume uma posição formalista na filosofia da matemática. Neste contexto, o objetivo desse capítulo é dar uma resposta negativa a esta questão, defendendo a visão de que Wittgenstein sempre pensou que não há separação lógica entre a matemática e suas aplicações. Nakano se concentra nos comentários de Wittgenstein sobre a matemática durante seu período intermediário, porque é nesse período que uma leitura formalista de seus escritos é mais tentadora. Por fim, o autor procura demonstrar como a ideia wittgensteineana de autonomia da aritmética não deve ser comparada à ideia formalista de autonomia, segundo a qual um cálculo é separado de suas aplicações. Na verdade, conclui o autor, a aritmética é autônoma, segundo Wittgenstein, porque garante sua própria aplicabilidade, proporcionando sua própria *raison d'être*.

No último capítulo, “Causalidade indeterminística e fenômenos quânticos”, Mayra Moreira Costa defende um modelo causal aplicável ao domínio indeterminístico da física quântica. Para a autora, o problema da aplicação da causalidade no domínio da mecânica quântica gera a dificuldade para explicar relações causais previstos pelo formalismo da mecânica quântica recorrendo às teoria e modelos causais clássicos. Nesse capítulo, para desenvolver seu argumento, Costa aborda o fenômeno do contexto experimental do tipo EPR-Bohm segundo o qual medições de propriedades de duas entidades quânticas espacialmente separadas apresentam uma anticorrelação perfeita entre eles e a dependência estatística entre elas não é anulada por uma causa comum, como ocorre no domínio clássico. Para fundamentar seu argumento, Costa propõe uma

generalização do Princípio de Causa Comum de Hans Reichenbach e uma reformulação do seu Critério de Transmissão de Marca, conjuntamente com uma reformulação do modelo de Bifurcação Causal Interativa de Wesley Salmon e da sua teoria dos processos causais. O modelo de Bifurcação Causal Conjuntiva de Reichenbach é usado para explicar os fenômenos clássicos, residindo na diferença entre os dois tipos de bifurcação causal a chave para compreendermos a distinção entre probabilidade e causalidade na física clássica e na física quântica.

Os textos premiados demonstram uma invulgar maturidade, tanto no domínio técnico dos procedimentos e temas peculiares à filosofia analítica, como na busca de formular e resolver problemas – isso que está na alma de todo filósofo filiado a essa tradição.

O Prêmio da SBFA e o presente livro expressam, assim, o vigor da criação filosófica brasileira como resultado da intensa atividade da comunidade de filósofos analíticos no Brasil.

Waldomiro J. Silva Filho
(Organizador)

SUMÁRIO

Apresentação: o movimento da filosofia analítica no Brasil hoje <i>Waldomiro J. Silva Filho</i>	07
Porchat, neopirronismo e a filosofia analítica (Discurso de Homenagem a Oswaldo Porchat Pereira) <i>Plínio Junqueira Smith</i>	12
Strong externalism about reasons <i>Rafael G. Vogelmann</i>	39
Color terms and perception <i>Raquel Krempel</i>	64
Wittgenstein, formalism and symbolic mathematics <i>Anderson Luis Nakano</i>	82
Causalidade indeterminística e fenômenos quânticos <i>Mayra Moreira Costa</i>	105

PORCHAT, NEOPIRRONISMO E A FILOSOFIA ANALÍTICA

DISCURSO DE HOMENAGEM A OSVALDO PORCHAT PEREIRA

PLÍNIO JUNQUEIRA SMITH
UNIFESP, CNPQ

1. Preservar a memória de um grande filósofo¹

Este texto serviu como base para uma apresentação oral em homenagem ao nosso querido e saudoso mestre, de quem, como vocês sabem, fui discípulo, não somente no sentido de ter estudado muitos anos sob a sua orientação, como também no de seguir, em linhas gerais, seu neopirronismo, ainda que o elaborando de outra perspectiva e com algumas divergências. Ele não tem como objetivo precípua apresentar de maneira sistemática as ideias de Porchat, embora inevitavelmente eu tenha de tocar em algumas delas com algum cuidado. Nos últimos anos, Porchat esteve muito doente, sofrendo de diversas e complicadas doenças. Por isso, sua capacidade de trabalho e sua locomoção foram muito afetadas. A consequência imediata é que, mesmo em São Paulo, onde Porchat residia, as novas gerações não tiveram a oportunidade de conhecê-lo em ação. É comum ouvir, mesmo de filósofos não tão jovens, que eles nunca conversaram com Porchat ou sequer o ouviram dar uma palestra; no melhor dos casos, o encontraram uma ou duas vezes de maneira superficial. Aos poucos, Porchat se foi convertendo numa espécie de mito na comunidade filosófica, alguém bastante conhecido e venerado por muitos, mas pouco presente e real para a maioria.

Porchat gostava de estar entre pessoas jovens, sua grande alegria era dar aulas e algumas vezes se dirigiu diretamente aos seus alunos, preocupado com a

¹ Eu gostaria de agradecer a Otávio Bueno, Paulo Faria, Roberto Horácio de Sá Pereira e a Waldomiro José da Silva Filho pelos comentários feitos a versões prévias deste texto e que me foram muito úteis para desenvolver ideias que, ou estavam confusas, ou sequer estavam presentes, mas deveriam estar.

formação filosófica que lhes estava sendo dada. O discurso de Porchat aos estudantes da USP em 1998, seguido depois por um bate-papo com estudantes aqui em Salvador, na UFBA, teve uma repercussão gigantesca Brasil afora. Assim, as ideias de Porchat sobre o ensino da filosofia se tornaram muito difundidas e influentes, mas seus textos filosóficos permaneceram pouco lidos. Uma razão para esse desconhecimento de sua filosofia era o fato, felizmente superado este ano, de que seu principal livro, *Rumo ao Ceticismo* (Porchat, 2006), se achava esgotado e até em sebos não era encontrado facilmente. Outra razão para isso é que, infelizmente, alguns importantes filósofos brasileiros parecem achar que somente seu livro sobre Aristóteles (Porchat, 2000), na verdade sua tese de doutorado defendida em 1967, tem uma qualidade excelente, mas seus artigos de reflexão pessoal seriam bem menos interessantes. Em certos meios, somente o Aristóteles de Porchat se converteu numa referência e num padrão de rigor, não as suas filosofias.

A envergadura de Porchat, no entanto, vai muito além de um excelente comentário a Aristóteles ou de ter ousado fazer uma autocritica na sua concepção da formação filosófica, colocando no seu devido lugar a importância da história da filosofia (importância essa que ele nunca desmereceu) e proclamando, em alto e bom som, que cada um deve dizer, por sua própria conta e risco, dando a cara a tapa, o que pensa de um problema filosófico, em vez de se limitar a produzir mais um comentário filosófico, porque, diante do conflito das filosofias, nada mais restaria a fazer, exceto tentar entender o “tempo lógico” ou a “ordem das razões” de um “sistema filosófico”. Onde o escândalo causado por seu belíssimo artigo – talvez o mais bonito de todos – *Prefácio a uma filosofia*. A meu ver, os seus textos de reflexão filosófica pessoal, a sua longa trajetória rumo ao ceticismo e o próprio neopirronismo que dela emerge paulatinamente constituem o essencial da sua contribuição à filosofia brasileira e, no entanto, são esses textos e essa trajetória que ficaram um pouco esquecidos. Mesmo o neopirronismo, como ele dizia (2006, p. 10) a “minha filosofia”, embora mais bem conhecido, ainda é insuficientemente compreendido e pouco debatido.

Esta é, portanto, uma oportunidade para preservar uma imagem um pouco mais fiel e começar a resgatar a contribuição do Porchat para as gerações mais novas, para colegas e amigos de outros estados que não puderam conhecer Porchat no auge de sua forma, para incentivá-los a voltar a ler seus textos, que merecem ainda análise detida e discussão minuciosa e rigorosa. Também é uma

oportunidade de dar a conhecer aos filósofos estrangeiros que nos honram com sua visita aquele que foi, para muitos de nós, certamente o maior filósofo brasileiro e talvez seja o único grande filósofo brasileiro até hoje. Pelo menos, como gostava de dizer Porchat, é o que me aparece.

2. Os papéis de Porchat para a filosofia analítica brasileira

Esta homenagem se faz em um congresso de filosofia analítica, promovido pela Sociedade Brasileira de Filosofia Analítica. Por que uma homenagem a Porchat em um congresso de filosofia analítica? A resposta mais simples é: porque Porchat foi decisivo para o florescimento da filosofia analítica no Brasil. E decisivo, em primeiro lugar, por uma questão institucional, já que Porchat criou o CLE – Centro de Lógica e Epistemologia, da UNICAMP. Como todos sabem, o CLE foi fundamental para a filosofia brasileira de um modo geral. Ao organizar colóquios, convidando vários dos melhores filósofos espalhados pelo Brasil, integrou a filosofia brasileira; ao convidar sistemática e regularmente inúmeros e importantes pensadores estrangeiros, inseriu e projetou a filosofia brasileira num cenário internacional. Porchat contava que mais de 70 filósofos estrangeiros vieram ao CLE nos primeiros 7 anos; Donald Davidson lá esteve, assim como John Searle, Kuno Lorenz, Gilles-Gaston Granger ou David Harrah. Um dos encontros da Sociedade Filosófica Ibero-Americana (SOFIA), o principal canal de interação da comunidade filosófica ibero-americana com os filósofos analíticos no mundo anglo-saxão, realizou-se em Campinas, com filósofos de toda a comunidade do continente americano. De fato, o CLE contou com vários professores estrangeiros, a começar com Ezequiel de Olaso, o grande amigo argentino de Porchat, que o ajudou a montar o CLE, tendo vivido por dois anos no Brasil, o lógico e filósofo da lógica Andrés Raggio, o filósofo da linguagem Eduardo Rabossi, Carlos Lungarzo, além de Harvey Brown, hoje professor na Universidade de Oxford, e Michel Ghins, hoje professor na Universidade Católica de Louvain, na Bélgica. O CLE criou três revistas importantes: a *Manuscrito*, os *Cadernos de História e Filosofia da Ciência*, e uma terceira de lógica que se fundiu com a *Journal of Non-Classical Logics*. Criou o primeiro curso de pós-graduação em lógica e filosofia da ciência da América Latina.

Mais especificamente, ao criar o CLE, Porchat colocou a pedra

fundamental da filosofia analítica no Brasil, dando-lhe não somente o seu impulso inicial, mas criando as condições de seu desenvolvimento sustentável, para usar uma expressão em moda. Foi graças ao CLE que a filosofia analítica pôs os pés em solo brasileiro de maneira firme e sólida, desenvolvendo-se de maneira frutífera e espalhando-se progressivamente pelo país. Se hoje estamos reunidos aqui, neste “V Congresso Brasileiro de Filosofia Analítica”, isso se deve, em proporção não pequena, embora já distante, a Oswaldo Porchat. A meu ver, essa já é razão bastante suficiente para homenageá-lo aqui, lembrando a sua triste morte, ocorrida em outubro do ano passado, aos 84 anos de idade. Acrescente-se a essa razão, como Porchat nos contava, em seu Discurso de Professor Emérito na UNICAMP, que o fato de que o CLE foi fundado no dia de 10 de setembro de 1975, portanto, há exatíssimos 43 anos. Não por acaso os deuses marcaram esta homenagem da Sociedade Brasileira de Filosofia Analítica para o dia de aniversário do CLE.

Institucionalmente, Porchat deu o pontapé inicial da filosofia analítica. Mas e quanto à sua obra? Era acaso Porchat um filósofo analítico? Pode parecer que não: ele foi, na década de 1960, um estruturalista formado na França, nos moldes de Martial Gueroult e Victor Goldschmidt, de quem foi discípulo; em 1968, decide abandonar a filosofia; já nas décadas de 1970 e 1980, foi um filósofo da visão comum do mundo, que pretendia promover filosoficamente as verdades comuns e resistir aos ataques de céticos e idealistas – seu apego à realidade e ao homem comum provinham, como ele mesmo confessava, de sua formação tomista e cristã (PF, p. 25-26); finalmente, de 1991 ao último de seus dias, foi um cético neopirrônico, diretamente nutrido por Sexto Empírico. Além disso, não me recorde de Porchat ter dito, uma vez sequer, que ele era um filósofo analítico (o que não quer dizer que ele nunca tenha dito isso). Ele se via primordialmente como um cético pirrônico, ou melhor, como um neopirrônico. Seu objetivo era atualizar o pirronismo antigo, tal como ele o entendia e tal como ele interpretava a obra de Sexto Empírico. Ele pretendia atualizar o pirronismo e via sua longa trajetória filosófica como tendo um tópos: o ceticismo.

Como um filósofo se referia a si mesmo e à sua própria filosofia, entretanto, não é decisivo para como nós vamos caracterizá-lo. É importante lembrar, nesse sentido, que a filosofia analítica nunca se definiu por uma doutrina filosófica, nem sequer por um método filosófico. A divergência doutrinária, e mesmo de métodos, sempre foi muito grande, de modo que há filosofias analíticas para todos os gostos em todas as áreas. Não há incompatibilidade alguma entre ser um

cético neopirrônico e ser um filósofo analítico. Prova disso é o neopirronismo do filósofo analítico Robert J. Fogelin (1994).

Tanto quanto posso ver, para ser um filósofo analítico basta pertencer a uma tradição em que certos autores e certas obras canônicas são referências quase obrigatórias e em relação aos quais um filósofo se situa quando elabora sua própria filosofia. E esses autores e essas obras definem o horizonte da assim chamada filosofia analítica. Toda a questão, portanto, é saber se Porchat, em sua longa trajetória rumo ao ceticismo, se insere nessa tradição da filosofia analítica, se certos autores e certas obras serviram-lhe de alicerce indispensável para as várias fases do seu pensamento. A esse respeito, a resposta parece inequívoca: Porchat é um filósofo analítico, pois esse longo percurso filosófico e essa renovação do pirronismo se fizeram, em grande parte, à luz da reflexão de vários dos principais filósofos analíticos: Moore, Ryle, Austin, Strawson, Quine, entre outros.

Antes de argumentar em favor dessa tese, cabe qualificá-la de duas maneiras. Primeira, é interessante notar que Porchat não se interessava muito pelas discussões mais detalhadas de filósofos analíticos menos importantes, embora ele possa ter lido e até usado alguns artigos ou livros menos conhecidos. Sua atenção se concentrava sobretudo nos grandes nomes da filosofia analítica. Segunda, a filosofia analítica se foi constituindo, a meu ver, cada vez mais como uma referência crucial para Porchat. Se, a princípio, o estruturalismo francês foi determinante, aos poucos Porchat foi se libertando desse estruturalismo e se aproximando da filosofia analítica. Parece-me que houve uma frequentação progressiva de autores da chamada filosofia analítica por parte de Porchat, sem a qual a sua filosofia da visão comum do mundo e o seu neopirronismo não teriam se constituído da forma como o foram.

Nesse sentido, a obra de Porchat se configura como o lugar privilegiado para uma discussão de filosofia analítica entre nós. Esse é o segundo papel que ele desempenhou para a filosofia analítica brasileira. Se, de um lado, incorporou a filosofia analítica em sua obra, de outro ele lega para nós, por meio de sua reflexão, ao menos duas filosofias analíticas, adaptando-as ao nosso contexto e à nossa língua. Se o CLE nos deu a forma, por assim dizer, da filosofia analítica do Brasil, a filosofia da visão comum do mundo e o neopirronismo nos fornecem a matéria.

Proponho, agora, uma rápida revisão de alguns aspectos importantes dos pensamentos de Porchat que se devem, em boa medida, a essa frequentação de

autores analíticos. O que se segue não é produto de uma investigação cuidadosa, mas apenas alguns lembretes, um tanto casuais, para orientar qualquer estudo sério nessa direção.

3. A filosofia analítica nas primeiras fases do pensamento de Porchat

Talvez se deva lembrar que, quando abandona a filosofia, em 1968, Porchat decide estudar lógica, porque a lógica não fala nada sobre o mundo, com Benson Mates, na Universidade da Califórnia, Berkeley. Quando volta ao Brasil, dá vários cursos de lógica matemática, que causaram impacto; segundo me contou Paulo Faria, o livro texto em que se baseavam suas aulas era o difícilíssimo e avançadíssimo *Mathematical Logic* de Joseph R. Shoenfield. Nesse período, e por um bom tempo depois, Porchat dá cursos de filosofia da ciência também, nos quais diversos filósofos analíticos são estudados. Em 1974, publica um artigo polêmico, “Contra o historicismo em teoria das ciências”, no qual defende uma forma de realismo científico e critica o historicismo. Esse realismo científico, diz Porchat, necessita de um realismo metafísico que o complemente (CHTC, p. 509). Notam-se nesse artigo os primeiros frutos da filosofia analítica no pensamento de Porchat.

Coube à sua filosofia da visão comum do mundo fornecer a base metafísica que faltava ao seu realismo científico. Nesse período, a primeira metade da década de 1970, o filósofo analítico mais importante de todos para Porchat talvez tenha sido George Edward Moore, com sua filosofia do senso comum. Certamente, Porchat marcou sua distância em relação a Moore, pois já tinha criticado o uso do senso comum como uma tentativa de resolver o problema do conflito das filosofias. Com efeito, diz Porchat, “minha postura em nada se assemelhava a uma promoção filosófica do senso comum, que alguns filósofos empreenderam” (CF, p. 34). Assim, Porchat entendia que a defesa do senso comum feita por Moore não resolvia o problema principal que as filosofias deveriam enfrentar. Além disso, a noção mesma de senso comum aparecia-lhe como ampla demais, pois o senso comum aceita uma diversidade de crenças falsas e supersticiosas, as quais o filósofo deve recusar. Segundo Porchat, “o senso comum varia muito no espaço e no tempo, no interior de uma comunidade ou de uma comunidade para outra” (FVCM, p. 49). Seria preciso, portanto, uma noção mais precisa para designar somente aquele núcleo duro de verdades inatacáveis

do senso comum. A esse núcleo duro, Porchat deu o nome de “visão comum do mundo” (FVCM, p. 51).

Embora Porchat tenha tentado se distanciar de Moore, o fato é que as duas filosofias são muito semelhantes. Jamais foi a intenção de Moore defender de maneira indiscriminada as crenças do senso comum. Moore (1993) elabora uma lista de ideias de senso comum, as quais teriam sido criticadas por um ou outro filósofo, mas, para ele, essas crenças são verdadeiras e nós sabemos isso com certeza. Essa lista de proposições de senso comum não é substancialmente diferente da lista elaborada por Porchat. Além disso, Moore reconhece que a expressão “senso comum” é muito vaga e que, tal como os filósofos a usam, ela pode conter “proposições... que não são verdadeiras” (1993, p. 119). Portanto, embora Moore insista em usar a expressão “senso comum”, parece que o que ele tinha em mente ao falar de certos tipos de proposições de senso comum que são verdadeiras não difere de maneira relevante do que Porchat chamava de “visão comum do mundo”. Uma vez, num curso de pós-graduação, após discutirmos um dos textos mais importantes de Moore, Porchat reconheceu essa semelhança e disse, um tanto surpreso consigo mesmo, que Moore tinha exercido uma influência mais importante sobre ele do que imaginava ou se lembrava. Como vimos, fora o tomismo abraçado por Porchat em seus anos de formação que lhe deu esse gosto pelo homem comum e pela realidade das coisas, mas foi Moore quem lhe forneceu o modelo filosófico atualizado desse realismo do homem comum.

Não por acaso, Porchat descreveu essa visão comum do mundo usando a clássica expressão de Peter F. Strawson: “metafísica descritiva” (FVCM, p. 70). Tratava-se, segundo Porchat, de explicitar a metafísica do homem comum (PF, p. 37). A filosofia da visão comum do mundo tomava a forma de uma metafísica descritiva, que descrevia a estrutura conceitual do pensamento do homem comum, recusando-se a especular sobre uma outra realidade ou a aprimorar nosso “esquema conceitual”. Lembro-me de quando fui a um seminário na casa do Porchat, em meados de 1980, e levava comigo o *Individuals*. Ao ver o livro em minhas mãos, Porchat elogiou o livro.

Até que ponto a metafísica descritiva de Strawson corresponde à filosofia da visão comum do mundo? Não está claro que Strawson estivesse reabilitando ao termo “metafísica” no mesmo sentido que Porchat o empregava – de verdades “metafísicas” no sentido tradicional, de verdades sobre as coisas em si mesmas, de um Absoluto, com maiúscula, como Porchat gostava de escrever. Parece, antes,

que Strawson tem em vista mais a generalidade de suas reflexões em relação a filósofos da linguagem ordinária, como Ryle e Austin, do que a uma dimensão metafísica propriamente dita, já que somente pretendia descrever nosso esquema conceitual, e não a estrutura do mundo.

No entanto, em favor de Porchat, cabe lembrar que o próprio Strawson se filia ao pensamento de Moore e não vê muita diferença entre a formulação sobre a realidade metafísica das coisas de Moore e sua formulação mais modesta de somente descrever a estrutura conceitual de nosso pensamento (Strawson, 1992, p. 32-33). A filosofia da visão comum do mundo de Porchat, então, nutriu-se desses dois grandes filósofos analíticos.

Se, na década de 1970, Porchat promoveu filosoficamente a visão comum do mundo, transformando as suas verdades em verdades metafísicas, na década seguinte ele procurou mostrar que essas verdades são incontestáveis, defendendo-as dos ataques desferidos pelos céticos (ou pelo que então ele via como ataques céticos). A própria formulação do problema e a estratégia empregada são claramente devedoras da filosofia analítica de um modo geral e de Moore em particular. Mas, num primeiro momento (SCC), o desenvolvimento detalhado dessa estratégia envolve outros filósofos que pertencem à tradição analítica ou até são inimigos da filosofia analítica, mas lidam com esta. Usando epistemólogos como Karl Popper (1974, 1981) e Imre Lakatos (1978, 1980), referindo-se a Ernest Gellner (1974, 1979), mas também a historiadores do ceticismo, como Richard Popkin (1979), Porchat propõe um modelo cético-cartesiano, o qual conteria o cerne desse ataque à visão comum do mundo. Seu intuito é defender a filosofia da visão comum do mundo desse ataque, o mais forte que a filosofia especulativa pode desferir contra o saber comum dos homens e suas verdades metafísicas.

Em seguida, a estratégia de Porchat foi a de identificar um elemento dogmático e especulativo na própria dúvida cética, ou seja, encontrar uma espécie de contradição interna ao ceticismo (CME). Ora, Ryle (1990) forneceu a Porchat os mais convincentes argumentos contra o ceticismo moderno, seja cartesiano, seja humeano. Outros filósofos, como Richard Rorty (1986), também criticavam a distinção mente-corpo, e Porchat estava bastante familiarizado com essa literatura. Nesse momento, Porchat achava que toda forma de ceticismo pressupunha algo como essa distinção; por ser questionável, um verdadeiro cético deveria rejeitá-la, isto é, um cético consistente deveria abandonar seu próprio ceticismo e abraçar o

realismo metafísico do homem comum a respeito do mundo, o qual gostaria, como um cético, de rejeitar. Muitos anos depois, Porchat ainda demonstrava grande interesse pela filosofia de Rorty, inclusive pelos seus textos morais.

4. O neopirronismo como um empirismo sem dogmas

Passemos para a década de 1990 com a emergência do neopirronismo. Para Porchat, Quine foi o maior filósofo do século XX, considerando *Word and Object* (1985) seu principal livro. Tenho a impressão de que Quine foi, para a elaboração do neopirronismo, o filósofo analítico mais importante para Porchat. Eis algumas razões para pensar assim.

O principal conceito do neopirronismo é o de fenômeno, em torno do qual se articula toda a doutrina neopirrônica, como o título mesmo do artigo inaugural dessa fase deixa claro: “Sobre o que aparece”. Porchat entendia que esse era um conceito obscuro no pirronismo antigo e para o qual não há uma boa tradução em nenhuma língua moderna. Seria preciso, então, “atualizá-lo”. A melhor tradução seria, no entender de Porchat, “experiência do mundo”. Ainda assim, era preciso explicar cuidadosamente o que seria essa experiência do mundo. Ora, das reflexões de Quine sobre frases observacionais e frases teóricas, Porchat retirou uma lição importante para a sua interpretação da noção sextiana do *phainómenon*: não haveria uma distinção de natureza entre os *phainómena* inteligíveis e os *phainómena* sensíveis, mas somente de grau, já que toda percepção sensível estaria, em algum grau, impregnada de termos teóricos (SA, p. 124). Embora, no final de sua vida, Porchat (NPSE, p. 302, nota 13) tenha voltado atrás nessa interpretação de Sexto Empírico e em sua doutrina segundo a qual só haveria um tipo de *phainómenon*, é certo que ele a sustentou por pelo menos duas décadas.

Tenho a impressão de que Quine também exerceu um papel importante para a concepção neopirrônica da verdade, pois Porchat não somente pensou a noção de fenômeno à luz da ideia quineana da continuidade entre frases observacionais e teóricas, como também pensou a relação entre frases observacionais e fenômenos (isto é, nossa experiência do mundo) como uma associação à qual fomos condicionados (VRC, p. 203). É porque diante de certas experiências fomos treinados a assentir a certas frases que se pode estabelecer uma conexão entre essas frases e essas experiências do mundo. Essa

correspondência entre o discurso e o mundo, indispensável na nossa noção comum de verdade, seria bem captada pelo behaviorismo quineano. (Uma certa inclinação pelo behaviorismo aparece cedo no pensamento de Porchat; PF, p. 31; FVCM, p. 43). Na próxima seção, tratarei da noção neopirrônica de verdade.

Além disso, Porchat entendia que o ceticismo pirrônico era uma forma de empirismo sem dogmas, expressão que tomou de empréstimo de Quine (1980, p. 42). Embora Porchat nunca tenha se alongado sobre esse tópico, parece-me certo que ele concebia o conhecimento humano e o conhecimento científico na esteira de Quine. Dado que o pirrônico antigo tinha uma visão um tanto ingênua da ciência e que a ciência moderna transformou profundamente o cenário do conhecimento científico, tornou-se uma questão premente para o pirrônico atualizado elaborar uma concepção de ciência que fosse compatível com a prática científica contemporânea (SA, p. 141-4). O empirismo cético é a chave para uma resposta a essa dificuldade (CA, p. 163). Aludindo repetidas vezes a Quine, Porchat elaborou a ideia de que o neopirronismo seria um “empirismo sem dogmas” (APSC, p. 266; ARMA, p. 286; EC, p. 319). Por exemplo, o neopirrônico deve atualizar o empirismo antigo, bastante próximo de um indutivismo de tipo humeano, à luz do método hipotético (EC, p. 320). Essa aproximação entre Quine e Popper, em torno do modelo hipotético-dedutivo e da continuidade entre a vida comum e a ciência como uma investigação empírica autoconsciente, Porchat a fez desde muito cedo. Após referir-se a Popper, Porchat diz que “também para Quine, a ciência não é mais do que o senso comum feito autoconsciente, valendo dizer o mesmo para a filosofia, já que ela não se distingue essencialmente da ciência pelo propósito ou método, mas apenas pelo seu escopo universalmente abrangente” (SCC, p. 85).

Devemos lembrar, ainda, da importante ideia de que o neopirrônico vive a vida comum como um homem comum. Essa ênfase em um terreno comum, pré-filosófico, no qual todos estamos inseridos, filósofos dogmáticos e filósofos céticos, filósofos e não filósofos, é um ingrediente fundamental do neopirronismo. Talvez seja importante notar uma guinada nessa noção de vida comum. Na fase da filosofia da visão comum do mundo, as referências analíticas de Porchat eram, como vimos, Moore e Strawson. Seguindo Moore, essas verdades comuns eram interpretadas como verdades metafísicas; seguindo Strawson, havia como que uma armação conceitual ou estrutura básica de conceitos, passível de descrição e explicitação filosóficas, “pois existe um núcleo massivo do pensamento humano que não tem história” (2006, p. 10). Porchat não somente deixou de lado a ideia

mooreana de que haveria verdades metafísicas incontestáveis no (núcleo do) senso comum, como também abandonou a ideia strawsoniana de que esse núcleo é permanente. “Nossa visão fenomênica do mundo se nos descobre como sujeita a uma permanente evolução” (SA, p. 138; CA, p. 172). Temos agora um todo sem fronteiras nítidas, na linha de Quine:

Podemos melhorar nosso esquema conceitual, nossa filosofia, pedaço a pedaço, enquanto continuamos a depender dele para apoio, mas não podemos nos separar dele e compará-lo objetivamente com uma realidade não conceitualizada. Portanto, não tem sentido, sugiro, investigar a correção absoluta de um esquema conceitual como um espelho da realidade (1980, p. 79).

Embora de maneira cautelosa, Porchat chega mesmo a falar de uma “essencial” contingência como uma das características mais conspícuas de nossa experiência fenomênica e a mesma armação básica de nossa visão do mundo não nos aparece como imune ao processo evolutivo” (SA, p. 138). Não surpreende, então, que Porchat ponha de lado a consagrada expressão de Strawson. “Não comportando asserções doutrinárias, não se conferindo uma dimensão cognitiva, ela recusa constituir-se como uma metafísica. Seria também prudente não caracterizá-la como uma ‘metafísica descritiva’” (SA, p. 137).

Para expressar essa nova concepção da vida comum, Porchat lança mão da metáfora quineana da “praça do mercado” (SA, p. 126; APSC, p. 267). A concepção de Quine cai como uma luva para essa “desdogmatização” (APSC, p. 272) da noção da visão comum do mundo (do núcleo duro do senso comum). Tudo quanto vimos sobre o empirismo é indispensável para compreender essa reviravolta de Porchat. O filósofo cético “vive a experiência do mundo e constrói e aperfeiçoa sua visão do mundo a partir dos recursos que lhe fornece a experiência” (APSC, p. 268). A ideia de que não há nenhuma crença totalmente imune à revisão e corrigimos nossa visão do mundo a partir da experiência, testando nossas hipóteses e melhorando-as à luz da ciência, é uma ideia central em Quine e no neopirronismo: não há um porto seguro, mas consertamos nosso barco em alto-mar, isto é, vamos modificando, corrigindo e melhorando nossa visão do mundo conforme ampliamos a experiência que temos do mundo. “A visão filosófica cética

do mundo não é mais que um saber empírico geral em permanente construção, evolução e aperfeiçoamento. Felizmente condenada a continuamente autocorrigir-se.” (APSC, p. 267). Em ainda outra alusão a Quine, Porchat reconhece que muitas de nossas crenças e verdades inscritas em nossa visão do mundo têm origem na “história da raça” (APSC, p. 266, 268), na experiência imemoriável perdida no tempo.

No entanto, ainda na esteira de Quine, a ciência dá continuidade a essa permanente reelaboração da visão fenomênica do mundo. “Porque reconhece, pelas razões que vimos, caber primordialmente à ciência a tarefa de investigar o mundo, o pirronismo enseja uma conjugação feliz entre ela e a filosofia, sem incorrer na ingenuidade positivista” (SA, p. 144). Não há mais fronteiras nítidas entre a visão comum do mundo e os resultados das ciências. Como dizia Quine, logo no início de *Palavra e objeto*, “a ciência é senso comum autoconsciente. E não se deve distinguir em pontos essenciais a filosofia, por sua vez, como um esforço de tornar as coisas mais claras, do propósito e do método da ciência, boa e má” (1985, p. 3-4). Há somente um todo articulado que se defronta com a experiência. Porchat fornece, como um exemplo histórico, a mudança da crença de que a terra é o centro do universo para a ideia de que a Terra gira em torno do sol (SA, p. 138). As ciências “não vão ser mais que a sistematização e metodização dos procedimentos do homem comum” (VRC, p. 197).

Portanto, não me parece exagero algum dizer que Quine foi de longe o filósofo analítico sobre o qual Porchat refletiu mais longamente e do qual incorporou mais elementos em seu neopirronismo. Com efeito, tal é a proximidade que Porchat vê entre Quine e o ceticismo, que ele chega mesmo a dizer que Quine “é bem mais cético” do que estaria disposto a aceitar (EC, p. 322).

Não se deve imaginar, contudo, que o neopirronismo não toma distância do pensamento quineano. Por certo, muitas doutrinas defendidas por Quine seriam recusadas pelo neopirrônico. “É manifesta a necessidade de o pensamento neopirrônico trabalhar sobre a problemática do empirismo contemporâneo, não menor que a necessidade de os teóricos do empirismo refletirem sobre a contribuição que a visão cética pirrônica do mundo pode oferecer-lhes” (EC, p. 322). É como se, dados o espírito cético da filosofia de Quine e o desconhecimento do ceticismo histórico por parte de Quine, este tivesse inadvertidamente ido além do que seu ceticismo lhe permitiria e adentrado o terreno dogmático. Nesse sentido, se tivesse consciência do seu ceticismo, Quine teria se corrigido em

diversos pontos. Porchat, infelizmente, não nos diz quais seriam as doutrinas quineanas incompatíveis com um empirismo sem dogmas. Mais adiante, na seção 7, voltarei à ideia de que o neopirronismo pode trazer benefícios para a filosofia analítica em geral.

5. A correspondência entre discurso e mundo

O artigo de Porchat “Verdade, realismo, ceticismo”, de 1995, revela a forte presença da filosofia analítica de uma maneira diferente da dos demais, porque não se trata de usar os filósofos analíticos para a sua reflexão, mas de entrar numa discussão específica travada pelos filósofos analíticos. Porchat se interessou pela questão da definição da verdade, pelas assim chamadas teorias da verdade, pela discussão que filósofos analíticos fizeram da noção de verdade. Ele se ateve detidamente ao debate entre Strawson (2017) e John L. Austin (1979), como também às contribuições de Davidson (1990) e Putnam (1981) sobre o assunto; ademais, dedicou-se com afinco aos trabalhos formais de Alfred Tarski (2007) e sua definição semântica da verdade, tendo discutido a sua interpretação com Newton da Costa.

De um modo geral, Porchat julgou que as críticas contemporâneas da tradicional teoria correspondencialista da verdade retomavam, sob nova roupagem, a velha crítica pirrônica à noção dogmática de verdade (VRC, p. 179). A concepção da verdade como correspondência sempre esteve ligada ao realismo metafísico. É de Putnam, que cita Hartry Field, que Porchat tira sua definição de realismo metafísico (VRC, p. 175, nota 2). Somente com Hume e Kant tal ligação teria começado a romper-se. Nesse sentido, muitas das teorias da verdade que encontram defensores entre os filósofos analíticos seriam teorias céticas, como a coerentista, a pragmatista ou a deflacionista. Haveria como que uma afinidade entre um defensor de alguma dessas teorias e o ceticismo: em ambos os casos, trata-se de rejeitar a concepção da verdade entendida como a correspondência (uma espécie de isomorfismo) entre o nosso pensamento (uma entidade mental) e a realidade (em si mesma). Assim, aos olhos de Porchat, muitas das teorias analíticas da verdade revelavam grande semelhança com o pirronismo antigo ao menos em seu aspecto crítico.

Ora, essa semelhança poderia sugerir que um neopirrônico deveria não

somente rejeitar a teoria correspondencialista da verdade, como também adotar uma dessas novas teorias (VRC, p. 180-2). Porchat, no entanto, não aderiu a nenhuma delas, porque, no seu entender, elas deixariam de lado a intuição comum da noção de verdade. E, embora tenha se interessado pela teoria de Tarski, não via nela uma descrição adequada do significado comum de verdade, faltando-lhe uma dimensão propriamente filosófica. Foi Austin quem forneceu a Porchat o auxílio fundamental para propor uma doutrina neopirrônica da verdade.

Em Austin (1979), Porchat encontrou uma definição da verdade na qual a ideia de correspondência seria preservada, sem, no entanto, incidir nos dogmatismos que a velha crítica pirrônica e a nova crítica dos coerentistas, pragmatistas e deflacionistas apontavam na tradicional teoria correspondencialista da verdade. Em outras palavras, ao propor sua definição de verdade, Austin em nenhum momento teria incorrido no erro de associá-la a um realismo metafísico. Por “realismo metafísico”, Porchat entendia a correspondência do que dizemos, não ao mundo de que temos experiência, mas a um mundo inacessível à experiência, um mundo “em si mesmo” ou “absolutamente independente”. Segundo Porchat, esse teria sido o grande pecado de Aristóteles, do qual Kant teria nos livrado. Embora defina a verdade por correspondência, essa explicação austiniana da verdade, por não se associar a esse “realismo metafísico”, não seria atingida pelos argumentos dos antigos céticos, bem como daqueles que atualizaram o arsenal pirrônico. Apoiado, então, em Austin, Porchat formulou sua concepção neopirrônica da verdade, em todos os aspectos relevantes distinta do que disse Sexto Empírico.

6. Wittgenstein

Muitos talvez estranhem a ausência até aqui de um nome fundamental para a filosofia analítica, o de Ludwig Wittgenstein. Embora preferisse Quine, Porchat conhecia bem sua importância para a filosofia contemporânea. É possível suspeitar que desde seu “O conflito das filosofias” a sombra de Wittgenstein se fez presente, seja numa menção à filosofia “como uma atividade metódica de elucidação e análise da linguagem comum ou científica” (CF, p. 17), seja no silêncio filosófico ao qual Porchat se confina (cf. PF, p. 31-32). Ambas as ideias estão presentes no *Tractatus Logico-Philosophicus* (Wittgenstein, 1999), o único livro de Wittgenstein

publicado em vida, tanto a ideia de que a filosofia é uma crítica da linguagem (1999, 4.0031), quanto a de que a filosofia correta é aquela que se recolhe ao silêncio (1999, 6.53). Mas, nesse momento, há somente alusão a Wittgenstein, se houver alusão.

Wittgenstein aparece explicitamente na fase da filosofia da visão comum do mundo com algum destaque. Vemos, por exemplo, Porchat se referir à metáfora da escada (PF, p. 32; SCC, p. 76-7, 85), também usada por Wittgenstein no contexto do silêncio filosófico (1999, 6.54). E Porchat não hesita em aludir ao famoso dito que encerra o *Tractatus* (“Sobre o que não se pode falar, deve-se calar” (1999, 7)), e que ele parafraseia assim: “se não se pode falar, calar-se é imperativo” (PF, p. 32). Tanto quanto Porchat àquela altura (PF, p. 31; FVCM, p. 42-43), Wittgenstein também rejeitava o ceticismo (1999, 6.51). Porchat parece coincidir com Wittgenstein em que já devemos conhecer, no ponto de partida, todos os fatos que precisamos conhecer (FVCM, p. 67). Essa insistência em trivialidades tem um caráter wittgensteiniano (FVCM, p. 58). Ao explicitar a concepção de mundo que se tem na visão comum do mundo, Porchat (FVCM, p. 45) se utiliza de uma formulação do *Tractatus*: “o mundo é a totalidade dos fatos” (1999, 1.1). Ora, todas essas semelhanças tornam inegável que Porchat incorporou ideias wittgensteinianas no silêncio da não filosofia e na filosofia da visão comum do mundo.

No entanto, nesses seus escritos iniciais, creio que Porchat sempre tomou distância de Wittgenstein. Tanto em “O conflito das filosofias”, como na filosofia da visão comum do mundo, tratava-se de rejeitar a concepção wittgensteiniana da filosofia como uma “análise ‘terapêutica’ da linguagem” (FVCM, p. 42) como não mais do que outra figura das filosofias dogmáticas a serem varridas pelo argumento do conflito das filosofias (CF, p. 17; FVCM, p. 42). O silêncio de Porchat não deve ser confundido com o silêncio de Wittgenstein, que, como uma lágrima, chega tarde demais (tomo essa metáfora de empréstimo a Paulo Leminski): “mas isso tinha de ser dito no começo, não no fim” (PF, p. 32). Daí que a metáfora da escada é rejeitada por Porchat: “Se não se tem a escada, não se pode nela subir. Nem se podem jogar fora escadas que não existem. A metáfora da escada sempre me pareceu enganosa” (PF, p. 32). Ademais, nesse momento, Porchat opõe-se frontalmente ao que vê como o “espírito idealista” que “tem impregnado nossa cultura”, com seu “culto abusivo da linguagem” (PF, p. 37). O positivismo lógico aparece novamente como um alvo de Porchat, por associar-se a um idealismo, ao

idealismo das filosofias contemporâneas da linguagem (FVCM, p. 64; cf. SCC, p. 84). Pode-se levantar a questão de se Porchat também incluía, ao lado do positivismo (e de tantas outras vertentes do pensamento contemporâneo), o *Tractatus* no rol dessas filosofias idealistas apoiadas em um privilégio desmedido da linguagem.

Em suma, para as duas fases iniciais de Porchat, o silêncio da não filosofia e a filosofia da visão comum do mundo, o Wittgenstein mais importante é aquele do *Tractatus*, bem como sua repercussão sobre os positivistas lógicos. Embora algumas aproximações sejam possíveis, resta que Porchat tinha mais em vista apontar divergências do que enveredar pelo mesmo caminho que o filósofo vienense.

E quanto ao neopirronismo de Porchat e a filosofia madura de Wittgenstein? Porchat leu com bastante atenção ao menos duas vezes as Investigações filosóficas de Wittgenstein (1968). A segunda foi mais ou menos na mesma época em que ele estava elaborando o seu neopirronismo; a primeira deve ter sido bem antes. Ele dispunha em sua biblioteca de vários livros sobre Wittgenstein. Uma vez, ele me emprestou o *Companion* do Garth Hallett (1977) e eu lhe emprestei um livro sobre Wittgenstein, acho que o *Insight and Illusion* de Peter Hacker (1986), o qual ele leu. Ele também tinha o *Companion* de Max Black para o *Tractatus*. Não eram muitos livros sobre Wittgenstein, mas o suficiente para demonstrar um claro interesse por sua filosofia.

Dada a desconfiança de que Wittgenstein tinha pela filosofia tradicional, dogmática e especulativa, não seria natural que Porchat tivesse usado Wittgenstein em seu favor? Afinal, reinou durante muito tempo na tradição analítica um questionamento acerca da legitimidade dos problemas filosóficos e, como Porchat reconhece, Wittgenstein desempenhou um papel importante nesse questionamento (CPPF, p. 222-3). Mais do que isso: essa afinidade entre boa parte da filosofia analítica contemporânea e o ceticismo pirrônico não se daria justamente por causa da filosofia de Wittgenstein? Ora, “os resultados de nossa reflexão cética são muito próximos dos que resultaram de diferentes análises filosóficas contemporâneas, todas elas empenhadas na desqualificação da legitimidade dos problemas tradicionais da filosofia” (CPPF, p. 237).

Além disso, Porchat faz uma série de referências positivas a Wittgenstein. Por exemplo, ao sustentar que a noção de vida comum é central para o neopirronismo, ele não alude somente à praça do mercado, de Quine, mas

também à expressão “forma de vida”, de Wittgenstein (por exemplo: 1997, 19) (SA, p. 126). A ideia mesma de que a filosofia é uma terapia, por exemplo, parece aproximar de maneira inequívoca o neopirrônico de Wittgenstein. Ora, o próprio Porchat cita frases famosas das Investigações filosóficas: “O pirronismo é, basicamente, uma crítica da linguagem e de seus mitos, ele luta para quebrar o feitiço que amarra os homens a uma linguagem em férias...” (SA, p. 141; cf. CA, p. 171). Ora, Wittgenstein já tinha dito no *Tractatus* que a filosofia é uma análise crítica da linguagem (1999, 4.0031), posição que ele mantém nas *Investigações Filosóficas*, quando diz que “a filosofia é uma luta contra o enfeitiçamento do nosso entendimento por meio de nossa linguagem” (1997, 109). Ainda nesse mesmo aforisma, Wittgenstein diz que “os problemas são dissolvidos, não produzindo nova informação, mas por meio da reorganização do já há muito conhecido” (1997, 109). A outra referência a Wittgenstein é: “Os problemas filosóficos começam quando a linguagem sai de férias...” (1997, 38). Porchat continua a dizer, em consonância com Wittgenstein, que “em filosofia tem muita importância lembrar as coisas que todo mundo sabe” (SA, p. 127; cf. VRC, p. 202). Também é digna de nota a reavaliação que Porchat fez da metáfora da escada. Se, a princípio, esta sempre lhe pareceu enganosa (PF, p. 31), posteriormente Porchat foi capaz de captar o seu sentido e perceber a sua força (AL, p. 334). Finalmente, Porchat diz que o homem comum está igualmente distante do realismo e do idealismo (VRC, p. 216), ecoando outra passagem famosa de Wittgenstein (1984, p. 48).

Embora tivesse grande admiração por Wittgenstein, Porchat não sentia nenhum fascínio especial por sua filosofia, a qual considerava dogmática, um tanto como a denúncia neopositivista da metafísica, por rejeitar os problemas filosóficos como pseudo-problemas. O neopirronismo não seguirá por esse caminho: embora denuncie os problemas filosóficos dogmáticos, muitos problemas filosóficos serão tidos como legítimos, enquanto outros problemas, embora dogmáticos, poderão ser reinterpretados como problemas fenomênicos. Na verdade, segundo Porchat, muitos problemas filosóficos legítimos são transformados, nas mãos de filósofos dogmáticos, em problemas filosóficos dogmáticos e, portanto, ilegítimos. Ora, é preciso revertê-los à sua formulação original e legítima. Do ponto de vista neopirrônico, há inúmeros problemas legítimos e aos quais o neopirrônico pode dar uma resposta. O problema da definição da verdade, que vimos na seção anterior, é um deles.

Voltemos a como Porchat interpretava a filosofia do segundo Wittgenstein.

Vimos como, em seus primeiros artigos, Porchat associava Wittgenstein aos positivistas lógicos, mesmo que não identificasse sua filosofia como positivista. Embora Porchat perceba uma diferença importante entre o segundo e o primeiro Wittgenstein, a ponto de lhe abrir uma exceção (CPPF, p. 238) ou, pelo menos, conceder-lhe o benefício da dúvida (CPPF, p. 227, nota 15), creio que ele nunca, de fato, pensou que não havia algum resquício de dogmatismo no pensamento de Wittgenstein. Em particular, Wittgenstein teria assumido uma postura próxima ao dogmatismo negativo: “E, quando Wittgenstein rejeita como pseudoproblema os problemas filosóficos tradicionais, o estilo de seu discurso deixa-nos a impressão de que o filósofo pretende ter mostrado de modo definitivo e conclusivo o não-sentido daqueles problemas” (CPPF, p. 227). Se Wittgenstein não é, propriamente falando, um dogmático, ele estaria mais próximo daquela forma de ceticismo praticada pelos cétricos acadêmicos.

A meu ver, “o estilo de discurso” de Wittgenstein pouco ou nada tem a ver com o estilo de Porchat. Ele sempre preferiu filosofias mais sistemáticas, como a de Quine. E sempre entendeu que a filosofia é essencialmente argumentativa: “a filosofia dogmática argumenta, ela essencialmente argumenta” (CA, p. 148). Ora, o ceticismo contra-argumenta, a ponto de se pretender “o representante mais autêntico, o legatário mais fiel da proposta mais fundamental e mais básica da filosofia do ocidente” (APSC, p. 270; cf. ARMA, p. 287). Quando se examinam os argumentos cétricos e os procedimentos de Wittgenstein, nota-se que, embora existam todas as afinidades mencionadas acima, Porchat recorre basicamente aos velhos argumentos pirrônicos: desacordo entre as filosofias, os demais modos de Agripa, modos de Enesidemo, método da antinomia. Talvez Porchat pensasse que, por trás da denúncia de ilegitimidade dos problemas filosóficos e da falta de sentido do discurso filosófico, havia mais uma concepção dogmática da linguagem e do significado do que uma contra-argumentação cétrica.

Também não se pode ignorar que Wittgenstein traçava uma distinção fundamental entre sua análise linguística e as investigações empíricas sobre o mundo. Com respeito à análise empírica, ele propunha uma terapia, cuja finalidade era eminentemente negativa ou, no melhor dos casos, permitia uma visão geral mais clara sobre as coisas. Ora, Porchat propõe que a investigação neopirrônica não esteja limitada, como em Sexto Empírico e Wittgenstein, a uma terapia, mas que se converta numa exploração empírica do mundo fenomênico. Nesse sentido, Porchat se acha muito mais próximo, como já vimos, de Quine. Paul Grice e

Strawson (1989) escreveram um artigo famoso, “Em defesa de um dogma”, no qual eles argumentavam contra Quine, preservando a ideia (supostamente wittgensteiniana) de uma distinção entre um domínio a priori da filosofia, no qual nos limitamos a observações linguísticas sobre o que já sabemos, em contraste como o domínio empírico das ciências, quando se propõem teorias a respeito do mundo e se avança o conhecimento. Ora, nesse debate, Porchat parece situar-se do lado de Quine, pois o empirismo neopirrônico se acha livre de dogmas, portanto do dogma da analiticidade. O neopirronismo inclui, em sua visão cética do mundo, tanto a dimensão comum, como as dimensões científica e filosófica, sem traçar fronteiras nítidas, como parece fazer Wittgenstein.

Mas, apesar disso, creio que, no final das contas, Porchat via Wittgenstein muito mais como aliado do que inimigo, pois as afinidades são evidentes. E, no final da vida, Porchat passou a ter uma concepção muito ampla do ceticismo: qualquer filosofia que recusasse o Absoluto seria uma filosofia cética, o ceticismo não sendo mais do que isso, a recusa do Absoluto. Nesse sentido, uma filosofia que dispensasse não apenas a argumentação tipicamente cética, mas a própria suspensão do juízo, poderia ser considerada cética. “Por isso mesmo, não é importante o fato de que certos elementos da filosofia de Sexto sejam historicamente datados. É o caso da *epokhé* e do método das antinomias, ingredientes básicos e da maior importância na estruturação da argumentação sextiana, mas de caráter puramente instrumental. Um neopirrônico de nossos dias não tem de retomá-los, eles não integram o núcleo central da proposta pirrônica” (NPSE, p. 321, nota 31). Nem, portanto, teria de retomá-los Wittgenstein.

7. Filosofia e história da filosofia

Já aludi ao fato de que Porchat ficou famoso Brasil afora por causa de sua chamada para o filosofar. Porchat fez um *mea culpa*, por ter exagerado na importância da história da filosofia para a formação filosófica, como se somente fosse possível se tornar um bom filósofo se se fosse antes um bom historiador da filosofia. Essa concepção da formação filosófica tem origem no seu estruturalismo dos anos 1960. Ora, Porchat progressivamente se afastou desse estruturalismo. Primeiro, deixou de lado as teses metafísicas; nisso o ceticismo antigo teve um papel importante. Depois, Porchat passou a considerar que o método estrutural em

história da filosofia não somente não garantia uma verdade objetiva, como também não era o único método (embora fosse o primeiro).

A filosofia analítica certamente contribuiu para essas mudanças no pensamento de Porchat sobre as relações entre filosofia e história da filosofia. Para Porchat, a história da filosofia devia servir para o filosofar. Se ele estudou detidamente filósofos como Sexto Empírico, Montaigne, Descartes, Locke, Berkeley, Hume e Kant, foi porque ele entendia que, para desenvolver certas ideias, esse estudo era indispensável. Por exemplo, reler Kant foi fundamental para a separação entre o realismo metafísico e a teoria correspondencialista da verdade. Ainda mais importante, seus estudos sobre ceticismo antigo, mormente Sexto Empírico, forneceram a perspectiva a partir da qual elaborar o seu neopirronismo.

Não se deve ver aqui, contudo, um apagamento das fronteiras entre a filosofia (entendida como a reflexão pessoal) e a história da filosofia (entendida como a interpretação de um sistema filosófico dado). Porchat jamais confundiu essas duas atividades. Uma prova disso é a diferença entre seu artigo “Sobre o que aparece” (Porchat, 1991) e “Sobre a noção de *phainómenon* em Sexto Empírico” (Porchat, 2013). Embora suas reflexões pessoais o levassem a ler Sexto e sua interpretação de Sexto ajudasse na elaboração de seu neopirronismo, essas duas atividades não se confundiam. Algo similar ao que Strawson fez em seu *Individuals* (Strawson, 2006a), uma filosofia pessoal de estilo kantiano, e em seu *The Bounds of Sense* (Strawson, 2006b), uma interpretação de Kant, quando a mútua relevância da filosofia e da história da filosofia se torna patente. O que Porchat ressaltava é que a história da filosofia deve ser útil ao filosofar e que devemos estudar história da filosofia na medida de nossos interesses filosóficos.

Vale a pena lembrar que, à certa altura, Porchat decidiu que não daria mais cursos de história da filosofia, mas daria cursos em que os alunos seriam estimulados pensar por conta própria. Fui aluno, com Roberto Bolzani, Sara Albieri e alguns outros colegas, desse seu primeiro curso. Textos de filósofos serviam apenas como leitura sugerida para que cada um redigisse duas páginas por semana nas quais expressava sua própria opinião. Era uma dificuldade... Não tínhamos o que dizer... Precisávamos nos esforçar para descobrir nossas próprias ideias. Pouco tempo depois, Porchat usou o clássico livro de Barry Stroud (1984), *The Significance of Philosophical Scepticism*, como base para um curso similar. Porchat não apenas usava textos de filósofos analíticos para essas suas aulas,

mas sobretudo ele promovia um estilo analítico de fazer filosofia.

8. A atualidade do neopirronismo: uma filosofia analítica eminentemente cética

Se muitos filósofos analíticos ofereciam reflexões para Porchat atualizar o pirronismo, é porque, segundo Porchat, boa parte da filosofia analítica é cética sem o saber. Essa é uma ideia recorrente na fase neopirrônica de Porchat. Se se pode falar do “caráter moderno e contemporâneo da postura filosófica do pirronismo” é porque salta aos olhos “o caráter eminentemente cético de uma parte considerável do pensamento filosófico do século XX” (APSC, p. 268). Por isso, diz Porchat, “a postura pirrônica é extremamente atual, inteiramente adequada às necessidades intelectuais de nossos dias. Nossa época está cansada de verdades, dogmatismos e especulações” (SA, p. 144). Na vida comum, nas ciências e na filosofia, o neopirronismo corresponde aos nossos anseios.

Porchat dizia que a filosofia analítica exibiu uma profunda ignorância do ceticismo histórico, repetindo e perpetuando “bobices que se associam à conotação do termo ‘cético’ no folclore intelectual do senso comum ‘filosófico’ de nosso tempo” (APSC, p. 268; cf. SA, p. 122; VRC, p. 198). Em particular, atribui-se ao ceticismo pirrônico o que é próprio do assim chamado ceticismo moderno. Para Porchat, “o ceticismo moderno tem alguma culpa na disseminação dessas tolices” (APSC, p. 268). A seu ver, o que os filósofos analíticos chamam de ceticismo não se parece em nada com nenhuma forma histórica conhecida de ceticismo, “o cético” dos epistemólogos analíticos não passa de uma construção útil para esses filósofos. O erro consiste em crer que essa construção de em alguma medida se assemelha ao cético histórico; se alguém assim entender (e muitos parecem entender), essa ficção útil não passará de um espantalho.

Por outro lado, também dizia Porchat, ao defender ideias bem modestas, ao deixar de lado exigências excessivas, os filósofos analíticos estariam abraçando o ceticismo sem o saber. De fato, muitos filósofos analíticos pensam que os impasses céticos são gerados pelas exigências absurdamente altas impostas pelos céticos; uma vez rejeitadas essas exigências, não haveria mais por que ser cético. Mas, para Porchat, quem coloca essas exigências muito altas são os dogmáticos e os céticos somente apontam para as consequências desastrosas de quem faz

essas exigências. Porchat insistiu longamente na ideia de que, se conhecessem bem a história do ceticismo e se soubessem o que de fato disseram os céticos históricos, muitos filósofos analíticos poderiam tranquilamente reconhecer-se como céticos.

Vimos que, na fase da filosofia da visão comum do mundo, Porchat julgava, de maneira pessimista, que nossa cultura era dominada pelo idealismo (PF, p. 37). Na fase neopirrônica, Porchat oferece uma descrição mais otimista: nossa cultura filosófica é, em boa parte, cética (APSC, p. 268). Ele sempre insistiu na adequação de suas propostas aos dias que correm, à utilidade da postura cética, por exemplo, para o empirismo em filosofia da ciência (EC, p. 321-2) e para outras filosofias se livrarem de algum ranço dogmático que porventura guardem (APSC, p. 272). Mas, para além da filosofia, o ceticismo é uma crítica constante dos dogmas que permeiam a sociedade, porque o dogmatismo é insidioso, dissimulando-se e permanecendo oculto.

A tarefa cética de ‘desdogmatização’ do discurso nunca se pode dar por definitivamente concluída. Os pirrônicos antigos dedicaram muita energia ao desenvolvimento de uma dialética capaz de questionar o dogmatismo aberto das doutrinas filosóficas. Parece-me que um pirrônico contemporâneo poderia, antes, dedicar mais atenção à detecção dos dogmas escondidos. Escondidos nos discursos dos outros e, talvez, em seu próprio discurso (APSC, p. 272-3).

Essa função social do neopirronismo encerra aquele que foi o último artigo publicado por Porchat (estou deixando de lado a conferência “Meu ceticismo”, que também foi publicada). Eu gostaria de citá-la por inteiro, já que vivemos, se não uma época, ao menos um período de forte obscurantismo:

E, nos dias que correm, a denúncia do dogmatismo se impõe como obrigação moral a todos nós. Guerras de religião e de extermínio, genocídios, massacres, holocaustos, regimes totalitários e desumanos se originam de variadas causas de natureza econômica, social ou política. Mas eles não podem de nenhum modo prescindir

dos discursos dogmáticos de que intensamente se servem para assegurar o domínio das mentes e conquistar adeptos. Educar os seres humanos contra o pensamento dogmático pode eventualmente tornar-se uma arma importante para prevenir, parcialmente ao menos, o advento de tais desgraças. E eu acredito que o conhecimento e a discussão do antigo pirronismo nos ambientes universitários pode contribuir para um tal propósito. (NPSE, p. 322)

Isso, é claro, se ainda existir um ambiente universitário onde se possa estudar ceticismo e combater os dogmatismos vigentes na sociedade.

9. Conclusão

Em suma, não somente do ponto de vista institucional Porchat foi fundamental para a filosofia analítica no Brasil, como também do ponto de vista filosófico, já que manteve uma relação intensa com a filosofia analítica, dialogando incessantemente com ela e adaptando-a a nosso contexto filosófico. Se, de um lado, ele frequentou e usou os principais textos da filosofia analítica, seja para sua filosofia da visão comum do mundo, seja para o neopirronismo, de outro, ele interpretou a modéstia que vigorou longo tempo na filosofia analítica como uma forma velada de ceticismo, como uma forma de ceticismo que não se sabe cética. A meu ver, ainda carecemos de um estudo mais minucioso das dívidas de Porchat com a filosofia analítica. O que eu disse acima não passam de alguns apontamentos iniciais para um estudo desse tipo.

Além disso, sua obra nos lega uma série de discussões críticas, de perspectivas originais sobre a própria filosofia analítica, convidando-nos a refletir sobre suas teses: é boa parte da filosofia analítica uma forma de ceticismo? Serão os argumentos céticos dialéticos suficientes para minar o projeto filosófico tradicional? Existem argumentos empíricos que conduzem ao ceticismo? É possível separar a noção correspondencialista de verdade de uma teoria metafísica da realidade? Há conhecimento que não seja absoluto? São os

problemas filosóficos legítimos, quando tratados empiricamente, e ilegítimos, quando discutidos de maneira dogmática? A percepção depende fundamentalmente de nossos conceitos? Essas são algumas das questões que Porchat levantou e que ainda estão aí, conosco, como também estão conosco as respostas dadas por Porchat: é possível uma teoria da verdade como correspondência que não seja metafísica; toda percepção pode ser expressa por proposições; mesmo que a percepção dependa de nossos conceitos, não se segue uma forma de idealismo. Cabe a nós retomar essas e, discutindo criticamente as respostas dadas por Porchat, levar a discussão adiante. Como disse meu amigo Paulo Faria (2017):

Mas essa obra admirável também é, na radicalidade do esforço de *começar de novo*, o ponto de partida incontornável do que um dia, talvez, teremos o direito de chamar por esse nome tão corrompido, 'a filosofia brasileira'. Ao empreender o esforço de acertar contas com a formidável provocação que constitui a obra de Porchat, estamos começando a superar *praticamente* o estágio das preparações e dos prolegômenos que foram tudo o que tivemos além de "análise de texto". Tal é, para tantos de nós, o sentido do itinerário exemplar daquele que não hesitamos em chamar nosso mestre – que tenhamos chegado a poder dizer, com o orgulho e a gratidão com que cumpre dizê-lo: 'Grande mestre, aprendemos tua lição. Estamos prontos para dar o próximo passo'.

Referências Bibliográficas

- Porchat Pereira, Oswaldo (2000). *Ciência e dialética em Aristóteles*. São Paulo: Editora da UNESP.
- Porchat Pereira, Oswaldo (2006). *Rumo ao ceticismo*. São Paulo: Editora da UNESP.
- Porchat Pereira, Oswaldo (1974). (CHTC) "Contra o historicismo em teoria das ciências". *Revista de História*. São Paulo: USP.

- Porchat Pereira, Oswaldo (2013). "Sobre a noção de *phainómenon* em Sexto Empírico". *Analytica*.
- Austin, John L. 1979 [1950]. "Truth". Em: *Philosophical Papers*. Oxford: Oxford University Press.
- Black, Max. 1964. *A Companion to Wittgenstein's 'Tractatus'*. Cambridge: Cambridge University Press.
- Davidson, Donald. 1990 [1984]. *Inquiries into Truth and Interpretation*. Oxford: Clarendon Press.
- Faria, Paulo. 2017. "O legado de Porchat". ANPOF: Coluna da ANPOF. <http://anpof.org/portal/index.php/en/comunidade/coluna-anpof>
- Fogelin, Robert J. 1994. *Neo-Pyrrhonian Reflections on Knowledge and Justification*, Oxford: Oxford University Press.
- Gellner, Ernest. 1974. *Legitimation of Belief*. Cambridge: Cambridge University Press.
- Gellner, Ernest. 1979 [1959]. *Words and Things*. London: Routledge and Kegan Paul.
- Grice, Paul; Strawson, Peter F. (1989) "In Defense of a Dogma". Em: Grice, Paul. (1989) *Studies in the Way of Words*. London and Cambridge, Massachusetts: Harvard University Press.
- Hacker, Peter M. S. (1986) [1972] *Insight and Illusion*. Oxford: Clarendon Press.
- Hallett, Garth. 1977. *A Companion to Wittgenstein's 'Philosophical Investigation'*. Ithaca: Cornell University Press.
- Lakatos, Imre. 1978. *Philosophical Papers*, volume 1. Cambridge: Cambridge University Press.
- Lakatos, Imre. 1980. *Philosophical Papers*, volume 2. Cambridge: Cambridge University Press.
- Moore, George E. 1993. *G. E. Moore: Selected Writings*. Ed. Thomas Baldwin. Oxford: Routledge.
- Popkin, Richard. 1979. *The History of Scepticism from Erasmus to Spinoza*. 2nd edition. Berkeley and Los Angeles: University of California Press.
- Popper, Karl. 1974. *Conjectures and refutations*. London: Routledge and Kegan Paul.
- Popper, Karl. 1981. *Objective Knowledge*. Oxford: Clarendon Press.
- Putnam, Hilary. 1981. *Reason, Truth and History*. Cambridge: Cambridge University Press.
- Quine, Willard v. O. 1980 [1953]. *From a Logical Point of View*, Cambridge, Massachusetts: Harvard University Press.
- Quine, Willard v. O. 1985 [1960] *Word and Object*. Cambridge, Massachusetts: MIT Press.

- Rorty, Richard. 1986 [1980]. *Philosophy and the Mirror of Nature*. Oxford: Blackwell.
- Ryle, Gilbert. 1990 [1949]. *The Concept of Mind*. London: Penguin.
- Shoenfield, Joseph R. 2010 [1967]. *Mathematical Logic*. Boca Raton, New York, Abingdon: CRC Press.
- Strawson, Peter F. 2006b [1966]. *The Bounds of Sense: an Essay on Kant's 'Critique of Pure Reason'*. London and New York: Routledge.
- Strawson, Peter F. 2017 [1950] "Truth". Em: *Logico-Linguistical Papers*. Abingdon: Routledge.
- Strawson, Peter F. 1992. *Analysis and Metaphysics: an Introduction to Philosophy*, Oxford Oxford University Press.
- Strawson, Peter F. 2006a [1959]. *Individuals: An essay in descriptive metaphysics*, London and New York: Routledge.
- Stroud, Barry. 1984. *The Significance of Philosophical Scepticism*. Oxford: Clarendon Press.
- Tarski, Alfred. (2007) *A concepção semântica da verdade: textos clássicos de Tarski*. São Paulo: Editora da UNESP.
- Wittgenstein, Ludwig. (1968) [1953]. *Philosophical Investigations*. 3rd Edition. New York: The Macmillan Company.
- Wittgenstein, Ludwig. (1999) [1922]. *Tractatus Logico-Philosophicus*. USA: Dover (reprinted from London: *Tractatus Logico-Philosophicus*. Routledge and Kegan Paul, 1955).
- Wittgenstein, Ludwig. 1984 [1958]. *The Blue and the Brown Books*, Oxford: Blackwell.

Siglas

- AL: "O argumento da loucura" (2003) Em: Porchat (2006)
- APSC: "Ainda é preciso ser cético" (2001) Em: Porchat (2006)
- ARMA: "A autocritica da razão no mundo antigo" (2005) Em: Porchat (2006)
- CA: "Ceticismo e argumentação" (1993) Em: Porchat (2006)
- CF: "O conflito das filosofias" (1968) Em: Porchat (2006)
- CHTC: "Contra o historicismo em teoria das ciências" (1974)
- CME: "Ceticismo e mundo exterior" (1986) Em: Porchat (2006)
- CPPF: "O ceticismo pirrônico e os problemas filosóficos" (1996) Em: Porchat (2006)
- EC: "Empirismo e ceticismo" (2005) Em: Porchat (2006)
- FVCM: "A filosofia e a visão comum do mundo" (1976) Em: Porchat (2006)

NPSE: "Sobre a noção de *phainómenon* em Sexto Empírico" (2013)

PF: "Prefácio a uma filosofia" (1975) Em: Porchat (2006)

SA: "Sobre o que aparece" (1991) Em: Porchat (2006)

SCC: "Saber comum e ceticismo" (1986) Em: Porchat (2006)

VRC: "Verdade, realismo, ceticismo" (1995) Em: Porchat (2006)

STRONG EXTERNALISM ABOUT REASONS

RAFAEL G. VOGELMANN
UFRGS

1. Introduction

Moral demands apply to everyone. A corrupted, hate-filled, insensitive person is not allowed to be cruel and one who no longer cares for others is not thereby relieved of the duty to help someone in need. Further, the fact that a moral demand applies to an agent gives her a reason to comply with it. Cruel persons have a reason not to be cruel and impassive persons have a reason to care. These two intuitive ideas are captured by the following thesis:

(Moral Absolutism) For every agent α and every action Φ , if Φ -ing is morally required, then α is morally required to Φ .

And

(Moral Rationalism) For every agent α and every action Φ , if α is morally required to Φ , then α has a reason to Φ .¹

Taken together these theses entail that:

(Categorical) For every agent α and every action Φ , if Φ -ing is

¹ I will follow Schroeder (2007: 17-9) in assuming that reason statements express a relation that holds between a fact F , an agent α and an action or action-type Φ – F provides a reason for α to do Φ . We get statements of the form “ α has a reason to Φ ” when we existentially quantify over the fact-place in that relation. We get statements like “ F is a reason to Φ ” when we universally quantify over the agent-place. See also Scanlon (2014: 30-1).

morally required, then α has a reason to Φ .

Some *prima facie* plausible claims about rationality and its connection with motivation seem to provide an argument against Categorical. The first premise in the argument is the widely shared thesis that the fact that an agent has reason to act in a particular way entails that she would have some motivation to act that way if she was fully practically rational and aware of the relevant facts. It can be thus formulated:

(State Internalism)² For every agent α and every action Φ , if α has a reason to Φ , then α has a pro- Φ attitude³ or α is not fully informed⁴ or α is not fully rational.⁵

² This thesis is a form of Internalism because it claims that every reason for action bears a certain relation to a motivational fact – this is the schematic form of Reasons Internalism (Finlay and Schroeder, 2017: 2). I call it “State Internalism” because it links the fact that an agent has a reason to the fact that the agent has a certain psychological state (in conditions of full rationality and information). Finlay and Schroeder (2017: 2) point out that one could defend *Motivation Internalism*, according to which if one has a reason to Φ , one is motivated to Φ in conditions of full rationality and information. Given that the fact that an agent is motivated to Φ allow us to ascribe her a desire (in a suitably broad sense) to Φ (this supposition is shared by a number of philosophers, such as Nagel (1978: 29), McDowell (1978: 15), Smith (1994: 116), Dancy (2004: 85)), Motivation Internalism entails State Internalism. I will focus on the weaker thesis. Further, State Internalism should not be confused with *Judgment Internalism*, according to which if an agent *judges* she has reason to Φ , then she either has a pro- Φ attitude or is not fully rational.

³ I will use “pro- Φ attitude” to refer to any pro-attitude that can be a source of the agent’s motivation to perform action Φ or that can be ascribed to the agent on account of her being motivated to Φ . Examples of pro- Φ attitudes are the desire to Φ , the inclination to Φ , the attitude of having the performance of Φ as a goal and the state of wanting to Φ . I will often use “desire” in a very broad sense that is interchangeable with “pro- attitude”.

⁴ I will use “fully informed” in a sense in which full information is restricted to non-normative facts. The fact that an agent is fully informed in this sense does not entails that she is aware of such facts as expressed by “the fact that F provides me with a reason to do Φ ”. If that were the case, every fully informed agent would have accurate beliefs about their reasons to act and then State Internalism would be entailed by Judgment Internalism. In the sense in

The second premise claims that an agent is rational to the extent she exercises her reasoning capacities in conformity with certain procedural rules (for instance, one does not violate rules of logic, engages in correct instrumental reasoning and imposes coherence upon one's set of beliefs and attitudes, perhaps by engaging in exercises of reflexive equilibrium). According to this view one is fully rational if, and only if, one fully exercises these capacities and, therefore, an agent can be deemed less than fully rational for lacking a particular pro-attitude only if she would come to acquire that pro-attitude were she to fully exercise her capacities for practical reasoning in a procedurally correct manner. This idea is captured by the claim that:

(Accessibility) For every agent α and every action Φ , if the fact that α does not have a pro- Φ attitude and is fully informed entails that α is not fully rational, then a pro- Φ attitude is accessible to α via fully informed procedurally rational deliberation.⁶

The last premise is the very reasonable supposition that:

(Supposition) For some agent α and some action Φ , Φ -ing is morally required and a pro- Φ attitude is not accessible to α via fully informed procedurally rational deliberation.

which I understand State Internalism, denying it does not entail the rejection of Judgment Internalism. See Parfit (1997: 124) for the distinction between normatively significant non-normative facts and normative facts.

⁵ See Nagel (1978: 20), Williams (1981: 109), Darwall (1983: 81), Smith (1994: 62, 150 and 1995: 109), Korsgaard (1996: 317), Velleman (1996: 694), Shafer-Landau (2003:168), Wallace (2006: 43), Finlay (2009: 4), Bedke (2010: 39), Parfit (2011: 62, 111). Cf. Johnson (1999: 53). I discuss this formulation of the thesis in detail in section 2.

⁶ The idea that the fundamental rules of practical rationality are purely procedural is, of course, widely shared. A sophisticated version of this thesis is defended by Smith (1994: 155–161; 1995: 112–3). See also Williams (1981: 104–105) and Korsgaard (1996: chapter 1). I discuss the notion of procedurally rational deliberation in section 3.

It seems, for instance, that a completely depraved person can engage extensively in correct, fully informed, procedurally rational deliberation and still fail to develop a desire or an intention to act as morality requires her to act. Caligula could be perfectly coherent and an all-around procedurally rational deliberator.

State Internalism and Accessibility jointly entail that for every agent α and every action Φ , if α has a reason to Φ , then a pro- Φ attitude is accessible to α via fully informed procedurally rational deliberation. In conjunction with Supposition, this claim entails that, in some cases, an agent is morally required to Φ but has no reason to Φ – i.e., it entails that Categorical is false.

Let us call this the *rationality argument*. If there is no equivocation in the premises, then it is valid. And let us refer to the position that denies Categorical as *Restrictive Internalism*, for it restricts morality ambitions, either by restricting the universe of agents certain moral demands apply to or by restricting moral demands' capacity to provide reasons for action.

Several philosophers reacted to the *rationality argument* by rejecting either Accessibility or Supposition. Those who reject Supposition argue that for any agent there is a solid procedural deliberative route that leads to the development of a desire to perform any action that is morally required of her.⁷ Those who deny Accessibility hold that there are substantive requirements of rationality that demand one to present certain desires and, therefore, that one can be declared less than fully rational for lacking a particular desire even though it is not accessible via procedurally rational deliberation. This is Parfit's view (Parfit, 2011: 78). I will refer to the position that denies Accessibility as *Weak Externalism*.

State Internalism, on its turn, has for the most part been regarded as a platitude.⁸ I will argue that it is not. It is a substantive thesis that can actually conflict with the procedural conception of rationality embodied in Accessibility and that can be consistently denied. It is possible to hold on to Categorical even if we accept

⁷ Smith, for instance, accepts State Internalism (Smith, 1994: 62), Categorical (Smith, 1994: 62-3) and a procedural notion of rationality that amounts to Accessibility (Smith, 1994: 155-6) and tries to deny Supposition (Smith, 1994: 187-9). His rejection of Supposition, however, is little more than an optimistic bet grounded on the fact that historically we have advanced in the direction of moral agreement. For a compelling objection see Sobel (1999: 145-7).

⁸ It has been explicitly characterized as such by Smith (1994: 62, 150) and Finlay (2009: 4). See also Shafer-Landau (2003:168) and Wallace (2006: 18).

Accessibility and Supposition.⁹ I will refer to the position that denies State Internalism as *Strong Externalism*.

I will not, however, present any direct argument against State Internalism. Indeed, in section 2, I will distinguish State Internalism from similar thesis with which it has been confused and show that it is not vulnerable to popular counter-examples to those thesis. I will rather argue that (i) the *rationality argument* begs the questions whether certain reasons are fact-given (i.e., whether we may have a reason to perform an action simply because certain facts obtain)¹⁰ and that State Internalism, Accessibility and the claim that there are fact-given reasons form an inconsistent set of thesis; (ii) that considerations concerning rationality do provide some independent reason to hold on to Accessibility, for its rejection leads to troublesome claims concerning the rationality of agents that lack pro-moral attitudes, but (iii) provide no reason to hold on to State Internalism, for its rejection does not lead to problematic consequences. Accepting both Accessibility and the claim that there are fact-given reasons (therefore, rejecting State Internalism) commit us to the Aristotelian thesis according to which fully rational and informed agents may fail to respond to reasons they have due to flaws in their education or development as practical agents that render them unable to grasp those reasons, but that is not a troublesome consequence.¹¹ The conclusion to be drawn is that the

⁹ In section 3 I will provide some reasons to doubt the prospects of defusing the *rationality argument* by rejecting Accessibility. As for Supposition, I will simply concede it to the proposer of the *rationality argument*. It should be clear that the argumentative burden faced by those who aim at denying Supposition is huge. These philosophers take head on the challenge of showing that, contrary to appearances, fully informed agents that lack attitudes corresponding to moral demands are always less than fully procedurally rational. Korsgaard (1996, chap. 3 and 4; 2009, chap. 3 and 4) is probably the most straight forward philosopher in arguing for the rejection of Supposition. For objections to Korsgaard arguments see Gibbard (1999), Fitzpatrick (2005) and Enoch (2006). For another attempt at grounding moral reasons procedurally see Schroeder (2007: chap. 6 and 7).

¹⁰ Among contemporary philosophers, the most prominent supporters of this thesis are Parfit (1997: 128-130 and 2011: 78) and Scanlon (2014).

¹¹ The idea that we can hold on to a procedural conception of rationality and to the idea that some agents have reasons that are not procedurally accessible to them if we accept this

idea that there are fact-given moral reasons that apply to every agent cannot be refuted by appeal to considerations about rationality, not even if we admit that rationality is purely procedural and that in some cases no amount of fully informed, correct procedural reasoning is enough to get an agent to conform to a moral demand. Rather, the truth or falsity of State Internalism can only be established after the debate about the existence of fact-given reasons has been settled on other grounds.

2. State Internalism

“Internalism about reasons” has been used to refer to a number of different but closely related thesis concerning the relation between reasons and motivation. In this section I will distinguish State Internalism from similar thesis and show that it is not vulnerable to counter-examples to these other thesis.

Bernard Williams, in his seminal paper *Internal and External Reasons*, holds that “if something can be a reason for action, then it could be someone’s reason for acting on a particular occasion, and it would then figure in an explanation of that action” (Williams, 1981: 106). According to him, if α has a reason R to Φ , then α can do Φ for R . Combined with the claim that only something that can motivate and agent can explain her action (Williams, 1981: 107), this yields the thesis that Schroeder calls “Existence Internalism about Reasons”, according to which “if there is a reason for someone to do something, then it must be possible to motivate her to do it for that reason” (Schroeder, 2007: 7). State Internalism and Existence Internalism are similar but distinct thesis. The latter is vulnerable to the following counter-example (Schroeder, 2007: 165): Nate hates parties, but loves surprise parties thrown in his honor when they are successful (that is, when he is actually surprised by them); there is a surprise party waiting for Nate at home; therefore, he has a reason to go home; nevertheless, Nate cannot be motivated to go home by the reason that there is a surprise party waiting for him for as soon as he becomes aware of that fact it stops being a reason for him to go home. The thesis that if α has a reason R to Φ , then it must be possible for α to be motivated

Aristotelian thesis has already been put forward by McDowell (1978 and 1995). I subscribe to his view on this matter. Part of my goal in this paper is to show its appeal.

to Φ by believing that R (Finlay, 2009: 2) is vulnerable to the same counter-example. State Internalism, however, is a weaker thesis than Existence Internalism and Nate's case is no counter-example to it. Given that Nate has a reason to go home, State Internalism entails that either he has a desire to go home or he is not fully rational or not fully informed and, as it turns out, Nate is not fully informed, for his being unaware that there is a surprise party waiting for him is a condition for him to have a reason to go home in the first place.

Another thesis that can be easily confused with State Internalism is the following:

(Exemplar) If α has a reason to Φ , then α^+ (i.e., α 's fully rational and informed version) has a pro- Φ attitude.

This is how Johnson (1999: 58) formulates Williams' internalist thesis. Johnson objects to this thesis by pointing out that sometimes agents have a reason to perform an action exactly because they are *not* fully rational and informed and, thus, that they would not be motivated to act accordingly if they were fully rational and informed. Consider this case: a fine day you come to believe you are James Bond; Exemplar entails that you have no reason to see a psychiatrist because if you were fully rational and informed you would not have this abnormal belief and, therefore, no motivation to see a psychiatrist (Johnson, 1999: 61). Similarly, if you are lost, you have a good reason to consult a map, but, if you were fully informed, you would have no such reason and would not be motivated to consult a map, for you would, *ex hypothesi*, know your way around. State Internalism is not vulnerable to these counter-examples and, therefore, cannot be identified with Exemplar. If you believe that you are James Bond you have a reason to see a psychiatrist, but you are probably not motivated to see one. Given State Internalism, what follows is simply that you are either not fully informed or not fully rational. In this case, it seems that both alternatives are true.¹²

State Internalism does not entail, therefore, that if less-than-fully-rational-and-informed α has a reason to Φ , then α would acquire a pro- Φ attitude by informing herself and getting over her rational flaws. State Internalism entails only

¹² The same considerations apply to Smith's angry squash player example (Smith, 1995: 111).

that if α , flawed as she is, has a reason to Φ and no pro- Φ attitude, then α is *currently* less than fully rational or fully informed. That is perfectly compatible with the possibility that as α becomes a more rational and informed person her reasons change.

Finally, consider:

(Advice) If α has a reason to Φ , then α^+ (i.e., α 's fully rational and informed version) desires that α would do Φ .¹³

According to Advice, if I have a reason to see a psychiatrist, my idealized counterpart would want me to see a psychiatrist. This thesis avoid the counter-examples just discussed. But State Internalism is weaker than Advice as well. Consider the case of an agent that is such that she would loose any interest or concern with herself if she became fully procedurally rational and fully informed.¹⁴ Suppose, for instance, that a person is so constituted that the impact of becoming aware of all the horrors in the world would completely obliterate her concern for herself in such a way that no amount of procedurally rational deliberation could fix that. According to Advice, this agent has no practical reasons whatsoever. I am not sure this is a counter-example to Advice, for I am not sure such an agent is really possible. But suppose it is. Would it be a counter-example to State Internalism also? No. State Internalism is compatible with the possibility that an agent has a reason to perform an action even though she would not want herself to perform that action were she to become fully rational and informed.¹⁵

State Internalism is a not a strong thesis. All it claims is that if one has a reason to Φ , then one can be declared fully rational only on the condition that one presents a pro- Φ attitude (unless one ignores a relevant fact). In other words, it claims that, for a fully informed agent, responding to all the reasons she has is a condition of full rationality. It is no wonder that many have taken it to be a triviality. We should not, however, identify State Internalism with a truism such as:

¹³ This thesis is defended by Smith (1995: 110-1).

¹⁴ See Bedke (2010: 43) for a similar case in which the fully informed and rational agent looses any interest in her non-idealized version.

¹⁵ Notice, however, that it does not entail that such a case is possible. It is also perfectly compatible with Advice. The point here is merely that it is weaker than Advice.

(Truism) If α has a reason to Φ , then it would be good-from-the-standpoint-of-reason for α to have a pro- Φ attitude.¹⁶

State Internalism is stronger than Truism. The latter claims that if an agent has a reason to Φ , then she would have a pro- Φ attitude if she was as her reasons recommend or advise her to be. It includes no mention of rationality. State Internalism, on its turn, claims that, for fully informed agents, having the attitudes one's reasons recommend one to have is a condition of rationality. These are different claims. One could confuse them only given the assumption that being rational is nothing but being as one's reasons recommend one to be. But that is an unreasonable assumption and one that is presupposed to be false by State Internalism. This becomes evident when we consider cases such as that of the man at a cocktail party who wants to drink gin and tonic and believes the glass before him contains gin and tonic when it actually contains petrol. Surely the man has a reason not to drink from the glass. But, given his ignorance of what the glass actually contains, the fact that he has no inclination to put his glass down without taking a sip does not testify against his rationality. He is not as his reasons recommend him to be, but he is (as far as this episode alone is concerned) perfectly rational. State Internalism admits of exactly this possibility: one can be fully rational even though one fails to be as one's reasons recommend one to be – all it takes is for one to be unaware of some relevant fact.

3. Weak Externalism

¹⁶ See Pettit and Smith (2006: 167) and McDowell (2006: 177-8). Pettit and Smith identify Truism with the claim that, (Advice*) if α has a reason to Φ , then α^+ (i.e., α 's version that fully exercises the capacities that ensure that her desires conform to principles of reason) wants α to Φ . This claim is quite like Advice. It can only be identified with Advice, however, on the assumption that the capacities the full exercise of which qualifies one as fully rational are the same capacities the full exercise of which ensures that one has the desires required by reason. This is exactly the assumption we reject if we reject State Internalism while holding on to the other premises in the *rationality argument*.

If State Internalism is such a weak thesis, it may seem that we have a better shot at saving Categorical if we reject Accessibility. According to Accessibility, an agent can be declared less than fully rational for lacking a particular desire only if she could acquire the desire in question by engaging in fully informed, procedurally rational deliberation. This claim embodies a conception of practical rationality: a perfectly rational agent is one that fully exercises, in a correct manner, a number of deliberative capacities that are governed by procedural rules. Call this the “*procedural conception of rationality*”. We may take a particular rule to be a procedural rule if it imposes no substantive condition on rationality, i.e., it does not impose as a condition of full rationality the possession of a particular attitude. A rule according to which an agent must care for her own well-being in order to be fully rational is substantive. A rule according to which one must have a desire to adopt ready-to-hand, effective means to one’s ends in order to be fully rational is procedural.¹⁷ It is an open question what are the procedural rules the compliance with is a condition of full rationality, but the proposer of the *rationality argument* can be quite liberal in answering it as long as Supposition remains plausible. According to Williams (1981: 104-5), for instance, procedurally rational deliberation includes not only means-end reasoning, but more complex and flexible forms of reasoning such as the combination of ends by time-ordering, resolution of conflict among ends, the search for constitutive solutions that specify what amounts to the satisfaction of a given end and exercises of imagination (one must, for instance, fully imagine the relevant effects of acting in a particular way or fully imagine what it would be like to satisfy a particular desire one nurtures). Smith adds to that list the capacity to engage in exercises of reflective equilibrium (Smith 1994: 158-61) and Parfit mentions the need to avoid wishful thinking and to assess probabilities correctly (Parfit, 2011: 62).

In order to reject Accessibility one has to claim that there are substantive conditions of rationality that are not procedurally established. That is, one has to defend an alternative conception of rationality, according to which the fact that a fully informed agent lacks certain desires may indicate a rational flaw even if she has reasoned perfectly from a procedural point of view. Call this the “*substantive conception of rationality*”. Now, why should one be declared less than fully rational

¹⁷ See Parfit (1997: 101-2) for a distinction between substantive and procedural rules along this lines.

for lacking a particular pro- Φ attitude despite being fully procedurally rational? I can see no explanation other than this: the agent in this case is less than fully rational because she has a reason to Φ but fails to respond to it. This is just State Internalism. We get the substantive conception of rationality if we add that there are reasons we have simply because certain facts obtain, even if the corresponding pro-attitudes are not procedurally accessible to us. Let us call these “*fact-given reasons*”. The substantive conception of rationality is, thus, a conjunction of two theses: State Internalism and the claim that some reasons are fact-given.

With the substantive conception of rationality in view one can see that the *rationality argument* begs the question whether there are fact-given reasons: State Internalism and Accessibility embody distinct conceptions of rationality that would conflict if one did not assumed beforehand that there are no fact-given reasons. On the one hand, we find State Internalism plausible because we take “fully rational” to mean something like “fully procedurally rational *and* fully responsive to the awareness of the facts that provide one with reasons”.¹⁸ And if that is the sense of “fully rational” with which we operate, we should find Accessibility to be groundless unless we assume there are no fact-given reasons. If part of being rational is developing appropriate attitudes in response to the awareness of the relevant facts, why should the fact that a fully informed agent is less than fully rational because she lacks a particular attitude entail that she is guilty of some flaw in procedural deliberation? She could be fully procedurally rational and be less than fully rational simply because she is not responding to a fact-given reason. On the other hand, if you are willing to accept Accessibility and the procedural conception of rationality it embodies, why would you accept State Internalism? That is, if you take “fully rational” to mean “fully procedurally rational”, why would you accept that if a fully informed agent has a reason to Φ then she has a pro- Φ attitude given that she is fully rational? That would be obvious only if you were already presupposing that our reasons are restricted by the attitudes that are accessible to us via procedurally rational deliberation.¹⁹ The *rationality argument*, thus, either rests on an

¹⁸ The simpler formulation “fully procedurally rational and fully responsive to one’s reasons” will not do. According to State Internalism an agent can be perfectly rational while failing to respond to a reason she has simply because she is unaware of a relevant fact.

¹⁹ One could argue that State Internalism is established by independent considerations about the role of reasons in explaining actions: if one has a reason to do Φ then one must

equivocation in the meaning of “fully rational” or begs the question of whether some reasons are fact-given.²⁰

One could argue that the idea that there are fact-given reasons involves very strong metaphysical commitments, so much so that we have good reason to reject it. But that would be too hasty. First, realists about fact-given reasons argue that internalists about reasons (if they are not eliminativists about reasons) also commit to the truth of normative statements according to which an agent has a reason to behave in a particular way because a certain fact obtains (namely, the fact that there is a suitably procedural connection between the agent's current attitudes and the relevant motivational attitude).²¹ Supporters of fact-given reasons simply claim that one can have a reason to behave in a particular way because facts of another kind obtain. Second, even if realism about fact-given reasons has further troublesome metaphysical consequences that are avoided by internalists, a realist interpretation of attributions of fact-given reasons is not mandatory.

be able to act for that reason and one can only act for that reason if one can motivated to do Φ by engaging in correct deliberation; therefore, if one has a reason to Φ , then in circumstances of full information and rationality one should have a pro- Φ attitude. That is a possible interpretation of Williams' argument (1981). If State Internalism was independently established in this way it could be combined with Accessibility to yield an argument for the rejection of fact-given reasons. However, even if the proposed argument is valid, it is an attempt at grounding State Internalism on Existence Internalism, to which a counter-example was presented in section 2.

²⁰ It is noteworthy that the *rationality argument* begs the question by making assumptions about reasons rather than about conditions of rationality. By now it is a commonplace to claim that those who appeal to a procedural conception of rationality in arguing that our practical reasons are restricted by the set of pro-attitudes that is procedurally accessible to us beg the question against a more robust conception of rationality. See, for instance, Korsgaard (1986), Hooker (1987), Parfit (1997: 116), Pettit and Smith (2006: 151-2) and Finlay (2009: 4). But I am making a further claim: even if we concede Accessibility to be true, accepting both Accessibility and State Internalism begs the question whether there are fact-given reasons. This means that the *rationality argument* begs the question not only against the substantive conception of rationality but also against those who combine the procedural conception of rationality with the thesis that there are fact-given reasons (a possibility to which I shall return in the next section).

²¹ See, for instance, Parfit (1997: 129) and Scanlon (2014:3-7).

According to a non-cognitivist view, for instance, attributions of fact-given reasons do not aim at capturing facts, but rather express an attitude of some kind and, therefore, involve no onerous metaphysical commitments. By itself, Externalism (whether in its weak or strong version) has no implications concerning the nature of reason statements and can be combined with different conceptions that range from realist to deflationary views. Accordingly, the *rationality argument* aims at establishing Restrictive Internalism without making any assumptions about the nature of reasons statements. If it was successful, it should compel even supporters of non-cognitivism about reasons to reject Categorical (despite their very parsimonious ontology). But, given that it begs the question whether there are fact-given reasons, it does not.

That, however, does not mean that the *rationality argument* is done for. In order to avoid Restrictive Internalism we still need to reject one of its premises. If the internalist can show that rejecting any of the premises leads to troublesome consequences, that would provide us with some reason to hold on to them and, therefore, with an argument (even if not a decisive one) for the rejection of fact-given reasons that would be grounded on considerations about rationality. And, as it turns out, the rejection of Accessibility does have troublesome consequences.

Given State Internalism, any claim about the existence of a fact-given reason entails a claim about the existence of corresponding substantive condition of full rationality and, therefore, the denial of Accessibility. Parfit argues in exactly this way: we all have, for instance, a fact-given reason to avoid future agony, therefore fully informed agents are less than fully rational if they lack a desire to avoid future agony, even if they are fully procedurally rational (Parfit, 2011: 78). That, however, is not enough to take the sting off the *rationality argument*. In order to accomplish that one would have to show not only that having some desires is a substantive condition of rationality but that, for every morally required action Φ , having a pro- Φ attitude is a condition of rationality. Otherwise the argument can be reformulated in a way that preserves its full strength: all we have to do is substitute “rational deliberation” for “procedurally rational deliberation” in Accessibility and Supposition yielding Accessibility* and Supposition*. “Rational deliberation” here is to be understood as procedurally rational deliberation that starts from a set of pro-attitudes that includes all the pro-attitudes whose possession is a *basic* substantive condition of rationality. The possession of an attitude A is a *basic* substantive condition of rationality if having A is a condition of rationality and A cannot be

acquired via procedurally rational deliberation from a set of attitudes that includes all the attitudes whose possession is a condition of rationality except A and those attitudes whose possession is a condition of rationality only because they are procedurally accessible from a set of attitudes that includes A . Let us refer to the attitudes required by basic substantive conditions of rationality as “*basic attitudes*”. In order to reject Accessibility* one would, then, have to claim that a fully informed agent can be declared less than fully rational for lacking a pro- Φ attitude even though the attitude in question is neither a basic attitude nor is it procedurally accessible (given full information) from the set of attitudes that includes all basic attitudes. This is absurd – it amounts to the claim that a perfectly procedurally rational and fully informed agent can be declared irrational for lacking a desire even though having that desire is not a substantive condition of rationality. Accessibility* is, therefore, a triviality.

This shifts the disagreement towards Supposition*. The question becomes whether for every morally required action Φ , a pro- Φ attitude is accessible via procedurally rational deliberation starting from a set of pro-attitudes that includes every basic attitude. Supposition* becomes less and less plausible to the extent that we distance ourselves from Accessibility and accept more and more basic substantive conditions of rationality. If the only basic attitude is, say, a concern for one’s own well-being, then rejecting Supposition* is almost as formidable a challenge as rejecting Supposition. On the other hand, if we accept, for instance, some altruistic inclinations to be basic attitudes then the challenge becomes less stringent. Therefore, in order for the rejection of Accessibility to provide a way out of the *rationality argument* that does not collapse into an attempt to reject Supposition or a very similar thesis, we would have to include in the set of basic attitudes at least some morally charged attitudes. It is at this point that this response to the *rationality argument* starts to lose its appeal.

Suppose that an altruistic inclination A of some sort is a basic attitude. This means that an agent can be fully procedurally rational, fully informed and otherwise substantively rational but still lack A , and to this extent be declared less than fully rational. The trouble is to provide some ground for that accusation. By stipulation, the accusation cannot be grounded on the claim that the agent fails to acquire A due to errors in procedural reasoning. The claim that the agent is less than fully rational can only be grounded, therefore, on the fact that the agent fails to acquire A in response to the facts that give her a reason to be altruistic. But the question then

is: how did we established that reacting in this way to these facts is a condition of rationality? We cannot have established it via procedural reasoning, for then the fully procedurally rational agent would agree with us. We have no option but to claim that basic substantive conditions of rationality are brute conditions of rationality. But then it seems that we have simply elevated the allegiance to part (perhaps a very central part) of morality to the status of a substantive condition of full rationality. The problem is that we seem to get this result for nothing. Why should the fact that an agent fails, say, to react properly to considerations about the needs of others show her to be rationally flawed instead of immoral, insensible or selfish? Williams complaint seems in place here: "Someone who claims the constraints of morality are themselves built into the notion of what it is to be a rational deliberator cannot get that conclusion for nothing" (Williams, 1985: 37).

The claim, of course, is not conjured out of thin air: it is the conclusion of a *modus ponens* that takes State Internalism and a claim about the existence of a fact-given reason to be altruistic as premises. The problem, of course, is that one's *modus ponens* is another's *modus tollens* and in this case the burden of proof seems to fall to those who want to be able to say of a fully informed, perfectly intelligible, intelligent, coherent, articulated, reflective, imaginative and stable agent that she is rationally flawed for lacking altruistic attitudes. The sense in which "rational" is ordinarily used is strongly associated with these other qualities, so much so that if one insists in saying of an agent that presents all these qualities that she is not fully rational it becomes hard to understand what one means by "fully rational" if not "fully procedurally rational *and committed to the central tenets of morality*".²² Parfit suggest that in declaring an agent irrational we are declaring her to be foolish, stupid or senseless (Parfit, 2011: 36, 114). But, surely, the coherent egoist is not stupid nor senseless. And she may be foolish in the sense that she fails to care for what really matters, but that is just to declare her morally blind. This brings out the force of Williams' claim that, given State Internalism, the claim that an agent has a reason to do something even though a corresponding pro-attitude is not procedurally accessible to her is nothing but a *bluff* (Williams, 1981: 111) – it is something we say hoping it will get our interlocutor to behave the way we expect

²² See Strandberg (2017: 75-83) for a battery of arguments to the effect that the substantive conception of rationality is at odds with our ordinary usage of "rational".

him to behave but which we cannot really back up when challenged.²³

I do not hope to have shown Accessibility to be beyond dispute. My point is merely that a philosopher who wants to convince us that the non-altruistic perfectly procedurally rational agent is not fully rational faces a tremendous argumentative burden and, therefore, that we have some reason to hold on to Accessibility.

4. Strong Externalism and the Aristotelian Thesis

State Internalism, Accessibility and the thesis that there are fact-given reasons form an inconsistent set of thesis. Weak Externalism exhorts us to reject Accessibility but that lead us to declare coherent, reflective and stable agents less than fully rational. That provides us with some reason to accept Accessibility, and, if we do, we have to choose between denying that there are fact-given moral reasons (i.e., accepting Restrictive Internalism) or denying State Internalism (i.e., accepting Strong Externalism). In this section I will argue that (absent any independent reason to deny the existence of fact-given reasons) Strong Externalism has no problematic consequences that could provide a reason to hold on to State Internalism.

If we accept Categorical, Accessibility (and the procedural conception of rationality that comes with it) and the thesis that at least some moral reasons are fact-given what follows is that, in some cases, fully informed agents can have a moral reason to Φ , lack any pro- Φ attitude and still be fully rational. One could argue that such cases are unintelligible for there is no way to explain the agent's failure to respond to her reason. We usually demand an explanation of such failure and, further, we expect it to point to a flaw in the agent – it must show that there is something wrong with her. Part of the appeal of State Internalism comes from the fact that it guarantees that there will always be one such explanation: the agent is either irrational or ignorant. If not that, what could explain the agent's failure to react to her reason?

This objection, however, overlooks a possibility. If we accept the claim that there are fact-given moral reasons, then we commit to the idea that there are substantive patterns of moral inference that are not necessarily intelligible to every

²³ For a similar complaint see Foot (1972:310).

fully informed and fully procedurally rational agent. For instance, if in a particular situation the fact that a person needs help provides a fact-given reason to help her, then it is correct to say that we have a reason to help *because* she needs help. If that is the case, one can fail to respond to a reason one has because one fails to grasp the relation indicated by that “because”. And then we can explain the agent’s lack of response to her reason in a way that shows there is something wrong with her without showing her to be rationally flawed: she is morally blind – she simply cannot see that *because* another person needs help she has a reason to help.

What is distinctive of Strong Externalism is, then, the claim that the set of capacities the full exercise of which ensure that the agent’s desires conform to her reasons is not exhausted by the set of capacities the full exercise of which qualifies the agent as fully rational. These are different capacities and an agent exposes herself to different kinds of criticism depending on which of those capacities she fails to exercise properly: in one case, she is declared rationally flawed; on the other, she is declared morally blind.²⁴

But what explains an agent’s inability to grasp the relation indicated by that “because”? What is the difference between the morally sighted agent and the morally blind agent? Both may be fully procedurally rational and fully informed. The difference between them can only be that the morally sighted agent is disposed to pass from the awareness of certain facts to the conclusion that she has a reason to act in a particular way (and to develop appropriate pro-attitudes) while the morally blind is not. If we take this suggestion seriously, we commit to the idea that our capacity to grasp moral reasons rest on certain dispositions (to acquire certain normative convictions and to develop certain pro-attitudes in response to certain facts) in such a way that fully rational and informed agents may fail to grasp reasons they actually have because their dispositions were not properly shaped by the education they received and by the experiences to which they were exposed during their development as practical and moral agents. The idea that an agent’s capacity to grasp moral reasons ultimately rests on a particular configuration of her

²⁴ This kind of distinction is perfectly ordinary. In an exactly similar way we say of an agent that is capable of making logical inferences but incapable of engaging in a kind of procedural practical reasoning (say, of combining ends in such a way as to make their satisfaction more convenient and economical) that she is theoretically rational but less than fully practically rational.

dispositions should be familiar to readers of Aristotle – the idea that a proper upbringing (that inculcates proper habits and dispositions in the agent) is a necessary condition of moral understanding is characteristic of his approach to practical philosophy.

Supporters of State Internalism may find this Aristotelian idea counter-intuitive. They may claim there is no room to speak in these circumstances of morally sighted and morally blind agents, but only of agents prone to different reactions. If the morally sighted agent simply has a disposition to acquire certain normative convictions and to react in a particular way in light of certain facts that the agent accused of moral blindness happens to lack, there is no ground on which to claim that the former is capable of grasping a reason that the latter fails to understand. And if that is the case, we cannot explain an agent's failure to react to an actual reason by declaring her to be morally blind.²⁵

It is, however, an error to suppose that our capacity to grasp moral reasons cannot be ultimately grounded on certain dispositions. Even our grasp of reasons grounded in procedural reasoning ultimately rests on the fact that we share some dispositions, for our very grasp of the forms of procedural reasoning in question rests on certain inferential dispositions.²⁶

Consider a very simple inference of the kind one must grasp in order to be capable of procedurally rational deliberation – say, a *modus ponens*. How could you

²⁵ Notice that supporters of Weak Externalism face a very similar problem even though they accept State Internalism and can, therefore, always explain an agent's failure to react to her reasons in terms of irrationality. It follows from the combination of State Internalism with the claim that there are fact-given moral reasons that having the set of dispositions that allow one to grasp those reasons is a condition of full rationality. Supporters of Weak Externalism, therefore, face the challenge of explaining why the agents that lack the relevant dispositions are less than fully rational and not merely agents prone to different reactions. This shows Strong Externalism to be preferable to Weak Externalism: both theses make the same assumptions (namely, that there are fact-given moral reasons) and face the difficulty just presented (even if in slightly different forms), but Weak Externalism face additional difficulties (that are avoided by Strong Externalism) for it leads to the onerous claims about the rationality of agents that lack moral pro-attitudes discussed in the previous section.

²⁶ As will become apparent, the argument that follows is inspired by Wittgenstein's rule-following considerations, specially §§185-242 of the *Philosophical Investigations*.

teach someone with no disposition to infer that q from $p \rightarrow q$ and p how to make that inference? It seems clear that if someone does not grasp such a basic inferential rule the only way to teach her will be to guide her along a number of examples and exercises, calling attention to the relevant features of the examples and correcting her errors in the exercises, until she becomes capable of making the inference on her own. But if one insists that our understanding of such inferences is not dependent on our inferential dispositions, what can this understanding amount to? Well, we are aiming at teaching an inferential rule, so we could say that one's understanding of *modus ponens* consists in the grasp of a rule, a kind of algorithm, which can tell the pupil what to do or how to apply a concept independently of what his spontaneous reactions are. If our goal is to convey a rule, then it must be the case that we expect the pupil to somehow infer what the rule is from the multitude of examples, exercises and punctual corrections we exposed her to – one may be tempted here to say that we expect the pupil to *interpret* this collection of elucidations. Let us call this the *interpretative conception* of our teaching methods.

Now, it is not at all clear what “interpreting” and “grasping a rule” mean in this context. One way to understand the interpretative conception is this: to interpret the examples and elucidations put forward by the teacher is to formulate (in a particular vocabulary) a rule that accounts for all of them; and to grasp the rule is just to have that formulation at your disposal. If the teaching process is successful, the pupil should formulate to himself a rule that accounts for the examples he was presented to – something like “to find the conclusion of a *modus ponens*, identify the consequent of the conditional premise”. Thus formulated, this picture of the process of teaching and learning faces an insurmountable problem: if we take it seriously, any attempt at teaching something new to someone leads to an infinite regress. For how can having a formulation of the rule help the pupil make simple inferences? That could help someone who already has a basic understanding of logic and knows what a *modus ponens*, a conditional premise and a consequent is. But, since that is not the case with our pupil (which we can safely assume, for otherwise we would not be teaching logic but just clarifying the meaning of an unknown expression of an already mastered notion), how can she know how to proceed according to the rule “identify the consequent of the conditional premise”? If she reacted immediately to this reformulation of the rule it could be helpful, but then her understanding of it would still be anchored in her dispositions. We could try to elucidate the reformulation of the rule by offering yet another reformulations of it,

but that would be the first step in an infinite regress.²⁷ Any reformulation of a rule will be useless if our pupil has no previous grasp of the notions they employ. The regress could only come to an end if we got to a formulation of a rule the pupil already understood. On the current reading of the interpretative conception, however, we have no idea what it means to say that pupil understands the rule if it is not to say that she has at her disposal a reformulation of it.

Formulating a rule cannot lead to learning if the expressions mobilized in that formulation has to be taught in the same manner – at some point, our methods of teaching have to take another form. The interpretative conception thus comprehended cannot provide the complete picture of our modes of teaching and learning. We should conclude there is a way of grasping a rule which does not consist in substituting one expression of it for another. Not all learning can consist in the production of a reformulation of a rule, on pain of making actual learning impossible.

Someone committed to the interpretative conception would have to drop, then, the supposition that to interpret examples and elucidations and to grasp a rule is always to come to an explicit formulation of it in a particular vocabulary. One could, however, hold on to the claim that the goal of teaching is to convey a rule that can guide us in isolation from any particular subjective dispositions by claiming that this rule cannot be formulated in any vocabulary. Grasping a rule is not a matter of coming up with a reformulation of its expression, yet it is a matter of grasping something from which the pupil can somehow infer, for instance, what is the conclusion of a simple set of premises. But what could this be? Perhaps something like an inner voice telling him what to do? But surely the fact that a voice comes from within does not change the fact that what it says is presented in a particular vocabulary and that we have to understand its commands. Maybe it is an image in the mind of the pupil? But one has to know how to react to an image as much as to a verbal expression. One has to be taught how to follow a road sign, for instance. If our pupil, instead of coming up with a verbal expression of the rule we are trying to convey, comes up with a mental image, she still have to learn how to

²⁷ Compare this with Lewis Carroll's puzzle in *What the Tortoise said to Achilles*. The Tortoise manages to paralyze Achilles exactly because he comes to accept that the gap between the premises and the conclusion in a simple *modus ponens* has to be filled with a formulation of the inferential rule.

use it. And how could we teach her that if not by presenting her with the same examples and exercises we would use to teach one to make the inference?

One could insist that the understanding of the rule the pupil has acquired cannot be expressed anyway at all (verbally or in the form of an image). But how could we comprehend this? It is as if a bolt of lightning crossed the mind of the pupil each time he was faced with an expression of the rule or an opportunity for its application, and then he knew how to proceed. But if that is all we can say, the fact is that the pupil in this scenario has developed a disposition to react in a particular way – a disposition to an unnecessarily complex two-step reaction, but a disposition to a particular reaction nevertheless. For instance, if when we say that our pupil has grasped the rule we were trying to convey by means of examples of simple *modus ponens* inferences all we mean to say is that when confronted with a particular set of premises of the form $\{p \rightarrow q, p\}$ she will know that the conclusion to be drawn is q , then to say that our pupil has grasped the rule is simply to say that he has developed an inferential disposition. And if that is all we can say, then we should conclude that we cannot teach basic inferential rules without fostering in our pupils particular dispositions.

If that is true of *modus ponens*, it must be true of the more complex forms of reasoning on which Restrictive Internalism grounds every reason. The fact, therefore, that Strong Externalism entails that our grasp of fact-given reasons rests upon a particular set of inferential dispositions is no objection to it. The same is true of our capacity to grasp reasons grounded in procedural reasoning, for our very capacity to engage in the required kind of reasoning ultimately rests upon a particular set of inferential dispositions. If we cannot say that the morally sighted agent actually grasps a reason that the morally blind agent fails to grasp because that capacity is grounded in a particular disposition, then we cannot even say that the agent that is capable of following a simple *modus ponens* actually grasps a logical relation that an agent incapable of following the same inference fails to grasp. And surely we can say that, so we must admit that the shaping of our dispositions is not necessarily an arbitrary process but can be aimed at enabling us to grasp modes of reasoning and, if there are fact-given reasons, certain reasons that would otherwise remain incomprehensible to us.

We should conclude that (absent an independent argument to deny the existence of fact-given reasons) we have no reason to prefer Restrictive Internalism to Strong Externalism. Therefore, given that the *rationality argument* begs the

question whether there are fact-given reasons, it gives us no reason to deny Categorical, even we are willing to accept Accessibility and Supposition.

5. Conclusion

The *rationality argument* begs the question whether there are fact-given reasons. If we accept the claim that there are fact-given reasons, then we have to deny either Accessibility or State Internalism. Denying Accessibility leads to very strong claims about the rationality of agents that lack pro-moral attitudes and this gives us some reason to prefer Restrictive Internalism to Weak Externalism. Rejecting State Internalism, on the other hand, commit us to the Aristotelian thesis according to which our capacity to grasp moral reasons is grounded on dispositions to present certain reactions that have to be shaped during our moral education and our development as moral agents. That consequence, however, should not be troubling (as long as we have no independent reason to deny the existence of fact-given reasons) for the same is true of our grasp of the procedural forms of reasoning on which Restrictive Internalism grounds every reason. Considerations about rationality, therefore, give us no reason to prefer Restrictive Internalism to Strong Externalism.

None of this, of course, proves State Internalism to be false – it is rather revealed to be a substantive thesis whose truth (if we accept Accessibility) cannot be established prior to the settling of the debate about the existence of fact-given reasons. The main lesson of this paper is, therefore, this: the question whether we all have reason to behave morally cannot be settle by considerations about rationality, not even if we accept the idea that rationality is purely procedural and that in some cases no amount of correct procedural reasoning is enough to get an agent to develop moral pro-attitudes. Quite the other way around, actually: once these assumptions are in place, in order to determine whether responding to reasons in conditions of full information is a condition of full rationality (i.e., whether State Internalism is true) we have to determine whether fact-given reasons exist or not. Of course the existence of fact-given reasons can and has been challenged on

many grounds,²⁸ but it cannot be challenged by considerations about rationality.

References

- Bedke, M.S. (2010). Rationalist restrictions and external reasons. *Philosophical Studies*, 151(1), pp.39–57.
- Dancy, J. (2004). *Practical Reality*, New York: Oxford University Press.
- Darwall, S. (1983). *Impartial Reason*. Cornell University Press.
- Enoch, D. (2006). Agency, Shmagency: Why Normativity Won't Come From What is Constitutive of Action. *Philosophical Review*, 115(2), pp.169–198.
- Finlay, S. (2009). The Obscurity of Internal Reasons. *Philosophers' Imprint*, 9(7), pp.1–22.
- Finlay, S. & Schroeder, M. (2017). Reasons for Action: Internal vs. External. *Stanford Encyclopedia of Philosophy*.
- FitzPatrick, W.J. (2005). The Practical Turn in Ethical Theory: Korsgaard's Constructivism, Realism, and the Nature of Normativity. *Ethics*, 115(4), pp.651–691.
- Foot, F. (1972). Morality as a system of hypothetical imperatives. *Philosophical Review*, 81(3), pp.305–316.
- Gibbard, A. (1999). Morality as Consistency in Living: Korsgaard's Kantian Lectures. *Ethics*, 110(1), pp.140–164.
- Hooker, B. (1987). Williams' argument against external reasons. *Analysis*, 47(1), pp.42–44.
- Johnson, R. (1999). Internal Reasons and the Conditional Fallacy. *The Philosophical Quarterly*, 49(194), pp.53–71.
- Korsgaard, C. (1986). Skepticism about Practical Reason. *The Journal of Philosophy*, 83, pp.5–25.
- Korsgaard, C. (1996). *The Sources of Normativity*, Cambridge: Cambridge University Press.

²⁸ See Scanlon (2014) for a catalog of such objections and a defense of fact-given reasons.

- Korsgaard, C. (2009). *Self-Constitution*, New York: Oxford University Press.
- McDowell, J. (1978). Are Moral Requirements Hypothetical Imperatives? *Proceedings of the Aristotelian Society, Supplementary Volumes*, 52, pp. 13-42.
- McDowell, J. (1995). Might There Be External Reasons? In: J.E.J. Altham & Ross Harrison (eds.), *World, Mind and Ethics: Essays on the Ethical Philosophy of Bernard Williams*. Cambridge University Press.
- McDowell, J. (2006). Response to Philip Pettit and Michael Smith. In: C. MacDonald & G. MacDonald (eds.), *McDowell and his Critics*. Malden MA: Blackwell, pp. 170–79.
- Nagel, T. (1978). *The Possibility of Altruism*, Princeton: Princeton University Press.
- Parfit, D. (1997). Reasons and Motivation. *Proceedings of the Aristotelian Society*, 71, pp.99–130.
- Parfit, D. (2011). *On What Matters – Volume 1*, New York: Oxford University Press.
- Pettit, P. & Smith, M. (2006). External Reasons. In: C. MacDonald & G. MacDonald (eds.), *McDowell and his Critics*. Malden MA: Blackwell, pp. 142–70.
- Scanlon, T.M. (2014). *Being Realistic about Reasons*, Oxford University Press.
- Schroeder, J. (2007). *Slaves of the Passions*, New York: Oxford University Press.
- Shafer-Landau, R. (2003). *Moral Realism: A Defense*, Oxford: Oxford University Press.
- Smith, M. (1994). *The Moral Problem*, Oxford: Blackwell.
- Smith, M. (1995). Internal Reasons. *Philosophy and Phenomenological Research*, 55(1), p.109.
- Sobel, D., 1999. Do the desires of rational agents converge?. *Analysis*, 59(3), pp. 137-147.
- Strandberg, C. (2017). A Puzzle About Reasons and Rationality. *The Journal of Ethics*, 21(1), pp.63–88.
- Velleman, J.D. (1996). The Possibility of Practical Reason. *Ethics*, 106(4), pp.694–726.
- Wallace, R.J. (2006). *Normativity and the Will*, New York: Oxford University Press.
- Williams, B. (1981). Internal and External Reasons. In: *Moral Luck*, Cambridge: Cambridge University Press.
- Wittgenstein, L. (2009). *Philosophical Investigations*, Trad: G.E.M. Anscombe, P.M.S. Hacker & Joachim Schulte, Oxford: Blackwell.

COLOR TERMS AND PERCEPTION

RAQUEL KREMPEL
USP

1. Some of the history of the debate¹

How well does color terminology mirror the way we conceive and perceive colors? The relation between color terms and perception and categorization has been a topic of discussion at least since the 19th century, when, in 1858, William Gladstone published his *Studies on Homer and the Homeric age*. In volume 3, section 4, Gladstone, the soon to be prime minister of the United Kingdom, calls attention to the peculiar uses of color terms in the Homeric poems. He notes that there is overall little reference to colors in the poems, and that there is more reference to black and white than to any other color. He also observes that the same color term is sometimes used to describe objects that don't seem to have the same color. Homer says that both the sea and oxen are wine dark; he also describes twigs and honey as *chlōros*, which is usually translated as green. Most shockingly, there is no word describing the color blue. Not even the sky is ever called blue in the poems. As Gladstone notes, "here Homer had before him the most perfect example of blue. Yet he never once so describes the sky. His *ouranos* is starry, or broad, or great, or iron, or copper; but it is never blue." (Gladstone, 1858, p. 483). Faced with the odd use of color terms in the poems, Gladstone offers what seems to him a plausible explanation for this: that "the organ of colour and its impressions were but partly developed among the Greeks of the heroic age" (Gladstone, p. 488).

What seems to be behind Gladstone's conclusion that the Greeks didn't see colors as we do is the commonsensical assumption that the way we perceive and conceive the world is clearly expressed in language: that language is a faithful

¹ I would like to thank João Vergílio Cuter, Evan Keeling and the GEMF group. This research was supported by grant #2014/15037-0, São Paulo Research Foundation (FAPESP).

COLOR TERMS AND PERCEPTION

mirror of the mind. Modern English speakers see red, yellow, green, blue, orange, etc., and they express what they see by the words “red”, “yellow”, “green”, etc. The distinctions that are expressed in language are the distinctions that are made in thought and perception. To the extent that the language of the ancient Greeks doesn’t make the same distinctions that are made in English, how could they have possibly seen colors the way we see them?

Gladstone’s suggestion that the ancient Greeks had an “immature” and “undeveloped” color perception sparked much discussion about the relation between color terms and the perception of color. As Deutscher (2010) recounts, in a book where he recovers the history of this debate, Gladstone had some supporters. Lazarus Geiger, a German philologist, found similar oddities in other ancient texts, like the Indian Vedic poems. Geiger, following Gladstone, suggested that color perception had evolved in the past millennia. But, as Deutscher notes, it didn’t take long for this idea to receive strong criticism. It was argued that animals, including insects, discriminate colors. This suggests that the ancient Greeks must have been at least able to see colors (for how could color vision have evolved in insects but not in the ancient humans?). Some, like Grant Allen, called attention to the use of pigments in Ancient art, including blue pigment (which was created in Egypt long before the Homeric poems were written). Why would the ancient Egyptians have gone to the trouble of creating pigments if they couldn’t see colors the way we do? And finally, questionnaires were sent to several parts of the world, in order to investigate the color vocabulary and perception of speakers of indigenous languages. They wanted to find out whether indigenous people had color terms coextensive with the color terms that appear in European languages, and whether they could discriminate colors like the Europeans. They found that they had no problem discriminating colors, regardless of what their color terminology was like. In fact, many languages, like Homeric Greek, did not have a word for blue, but their speakers could still discriminate between, say, blue and green. For all these reasons, Allen concludes “that the want of colour-epithets in the Homeric poems is due to a defect of language rather than of perception” (Allen, 1878, p. 132).

We don’t, of course, need to go as far as Allen and say that languages that don’t, e.g., have a word for blue are defective. For what color words a language has might well be a reflection of cultural and environmental needs. But these are certainly very convincing arguments against the idea that the language one speaks

reflects the colors that one can see and the color discriminations that one is able to make. Since then, it has simply been taken for granted that all humans with normal color vision can see and discriminate approximately the same colors (Roberson and Hanley (2010) speak of 2 million hues). But even if we accept that we can all see that hue x is different from hue y in approximately the same way, and that language is not a perfect mirror of what goes on in perception, there are still some questions that we can ask about the relation between color terms and perception. For instance, does the language one speaks reflect how one conceives or categorizes colors, or do we all group colors into categories in the same way, despite differences in color vocabulary? Can the language one speaks somehow influence the way one perceives the similarity and dissimilarity between different colors – that is, does grouping two hues in the same linguistic category make them appear or be judged to be more similar than two hues that are equally spaced but that receive different color terms? And if so, how exactly does language exert that influence? These questions started to be tackled more recently, as we will see in the next section.

2. Universal color concepts

In a famous study, Berlin and Kay (1969) have suggested that color terminology evolves in predictable ways in different languages, and that we all form color categories around the same universal foci. They considered 20 different languages and noted that despite the fact that they differed in the number of basic color terms they had (ranging from 2 to 11), speakers of all the languages they tested placed the best examples of their color terms in predictable areas of the color spectrum. This showed, according to them, that languages don't cut the color spectrum in completely arbitrary ways. Berlin and Kay took this to mean that there is something about the biology of color perception that leads us all to categorize colors in similar ways (around the same focal colors). So the way people perceive and conceive colors should presumably not differ due to the language they speak, contrary to what Gladstone had suggested.

This point received stronger empirical support from Heider's (1972) investigations with speakers of Dani, a language spoken in New Guinea. The Dani language has only two basic color terms, which reflect an opposition between black

and white, or light and dark. Despite their color terminology being very different from English (which has eleven basic color terms), they behaved very similar to English speakers in memory tasks. For instance, just like English speakers, Dani speakers showed better memory for colors that are focal for English speakers (say, a good example of blue), than for colors that are not focal for English speakers (say a color that is on the boundary between blue and green). Heider took this to mean that some colors are universally recognized as focal. So Dani speakers presumably had the same color categories that English speakers have, around the same centers, despite their limited color vocabulary.

Among philosophers, a common view, accepted for instance by Jerry Fodor, is that sensory concepts are innate, color concepts being the typical example.² Fodor argues in fact for a stronger view, according to which most of our primitive concepts (concepts that are not structured around simpler concepts) are innate. By saying that primitive concepts are innate, Fodor doesn't mean to say that they were literally present at birth. What he means is that mere experience with properties in the world is enough to trigger the acquisition of primitive concepts. He thinks, though, that experience, being so varied, is not sufficient to explain the uniformity of the concepts we end up having. Given that experience itself doesn't seem sufficient to guarantee that concepts will be uniformly created across different people (which he is assuming is what happens), due to different individual histories, it seems that primitive concepts need to, in some sense, come from within. We come to the world predisposed to form the primitive concepts we do when exposed to certain properties in the world.

This general view about primitive concepts applies to the case of color concepts. Fodor believes that color concepts are acquired by mere experience with colors in the world. As he says, "we typically get the concept RED from (or, anyhow, on the occasion of) experiencing things *as red*" (1998, p. 131), and that "all that's required for us to get locked to *redness* is that red things should reliably seem to us as they do, in fact, reliably seem to the visually unimpaired" (1998, p. 142). His idea is that our color concepts are simply triggered by their instances in consequence of

² According to him, "on all standard theories the sensorium is an innately specified function from stimuli onto primitive concepts. The range of potentially available primitive-sensory concepts is thus fixed when you specify the biological endowment of the organism, as is the range of stimulations which can occasion the availability of those concepts." (1981, p. 276).

our biology; color concepts are not learned.³

If Fodor is right and all that is required for us to acquire a color concept is that we enter in causal contact with colors in the world, then it seems clear that the language we speak should have no impact on the way we conceive of colors.⁴ Color concepts are formed as a result of our experience with colors in the world. So we shouldn't expect variation in color concepts in people who have been acquainted with the same colors. If there still is linguistic variation in color terminology across languages, even assuming their speakers had equal experience with colors, then the difference is likely to be purely linguistic. Color concepts and perception would presumably be immune to the influence of language.

I have said so far that, unlike what some thought in the 19th century, the language one speaks doesn't reflect the color discriminations one is able to make. After Berlin and Kay's book, the idea that color concepts are centered around universal foci and determined by our biology gained popularity. The suggestion was not only that we can all make the same color discriminations, but that we all group colors together more or less in the same way – that is, we all conceive colors in the same way –, even though languages can vary in the number of basic color terms they have. If that is really the case, then it seems that people like Fodor may be right in saying that color concepts are innate. Presumably the language one speaks neither *reflects* how one conceives and perceives colors (for everyone with approximately the same experience with colors should conceive colors in the same way), nor does it cause any differences in the way one conceives and perceives colors.

However, some recent findings have challenged these views. There is now growing evidence in support of the idea that linguistic differences are not only terminological, but that the natural language one speaks has some influence on the way colors are conceived and discriminated. For instance, speakers of different

³ Fodor seems to be accepting realism about colors, that is, the view that colors are real properties of objects. But I'm not going to discuss this traditional philosophical problem of determining the nature of color properties.

⁴ I am here ignoring Fodor's view that monomorphemic words typically express primitive concepts. Given the variation in color terms across languages (which presumably express primitive concepts), one might find this view incompatible with the idea that primitive concepts, such as color concepts, are innate.

languages will show different categorical effects both in memory and in perception. In memory tasks, that effect appears when subjects are more accurate at recognizing a previously seen color when it belongs to a different category from the alternative they have to choose from. Categorical perception occurs, for instance, when one is faster at discriminating two colors when they belong to different categories than when they belong to the same linguistic category, despite their being all equally spaced according to a color metric. These effects have been found not to be universal, but are instead consistent with the color categories in the language a person speaks.

Assuming that the language one speaks can influence the way a person remembers and discriminates colors, a further question is *how* exactly language does that. Some studies on the categorical effect of color terms have suggested that linguistic representations interfere directly in color memory and discrimination. The idea is that words play an online role in perception. In the next section, I will review some of these findings. In section 4, I will criticize the conclusion that linguistic representations act directly on perception. I will suggest an alternative explanation, namely, that language acts indirectly on perceptual discriminations, through the shaping of color concepts (conceived as non-linguistic representations that might either belong to the visual system or be amodal, thus belonging to higher cognition).

3. Color terms and perception: the empirical evidence

There are now several studies that claim to show that linguistic representations somehow interfere in perceptual tasks. We can start by considering an older one, moving then to consider more recent studies. We will see that they differ in how they interpret the influence of language on color discrimination and memory: some say that language modulates perception directly, whereas others claim that language interferes in perceptual judgments or decisions. These differences won't matter for our discussion here. What is important for my purposes is that they all claim that *linguistic representations* have some role to play in color memory and discrimination.

In one experiment, Kay and Kempton (1984) compared English speakers with speakers of Tarahumara, an indigenous language spoken in Mexico that,

unlike English, has only one basic color term for both blue and green. In each trial, participants had to indicate which of three color chips, which were all in the green-blue region, was the most different from the other two. They found that “the English speaker judges [color] chip B to be more similar to A than to C because the blue-green boundary passes between B and C, even though B is perceptually closer to C than to A” (Kay and Kempton, 1984, p. 77). The same effect was not observed in speakers of Tarahumara. What they think explains this finding is that, in order to solve a difficult task, English speakers unconsciously use what Kay and Kempton call a “name strategy”. Because they find it difficult to say which color chip is perceptually more distant from the other two, they use the tactic of choosing the one that receives a different name. The authors take this reliance on color names to be unconscious because participants explicitly claimed not to be judging color similarity based on language and insisted that colors that crossed the blue/green boundary looked more different than colors that were more spaced out but that belonged to the same linguistic category. Since speakers of Tarahumara had at most one color name to rely on, their judgments of similarity were no different when the colors crossed the blue-green boundary.

Kay and Kempton believe that these differences in similarity judgments between English and Tarahumara speakers “are obviously not caused by an aberration of vision” (p. 75) in English speakers.⁵ Instead, they say that “there do appear to be incursions of linguistic categorization into apparently nonlinguistic processes of thinking, even incursions that result in judgments that differ from those made on a purely perceptual basis.” (p. 77) So they are adopting a distinction

⁵ In order to determine the “real” perceptual distance between the color chips they used, Kay and Kempton adopted a scale of psychological distance for colors whose unit “is the just noticeable difference (jnd), that is, the smallest physical difference in wavelength that can be detected by the human eye.” (p. 68) They then say that “it cannot be the case that the vision of English speakers is distorted in some way by the language they speak, because the discrimination distances that the Tarahumara faithfully reproduce on the subjective triads task were established on speakers of English.” (p. 72) However, one could object that while the scale based on just noticeable difference is one thing, the subjective impression of similarity between colors that are farther apart is something else. So one could still try to insist on the idea that the differences in judgments between English and Tarahumara reflect differences in perception. But I won’t push this point here.

between judgments that have a purely perceptual basis, and judgments that involve higher cognitive processes. What is happening, according to them, is that linguistic categories unconsciously interfere with English speakers' judgments about perceptual distances when they are presented with a difficult task. They designed a second experiment that blocked the name strategy and that forced similarity judgments to be made by taking into consideration perceptual information alone. They found that this blocking of the name strategy resulted in more accurate judgments from English speakers about the actual perceptual distances between colors.

More recently, others have shown that there are behavioral differences among speakers of languages whose color terminology is not coextensive. Roberson et al. (2000, 2005) have performed a series of experiments with Himba, Berinmo and English speakers, and similarly found differences in performance. Both Berinmo and Himba have only five basic color terms (English has eleven), and neither language marks the distinction between blue and green. One experiment was very similar to Kay and Kempton's, involving similarity judgments, and they got similar results. In another experiment (similar to one that Heider performed with Dani speakers), they tested the influence of language on memory. Roberson et al. (2000, p. 392; 2005, p. 399) showed speakers of Himba, Berinmo and English one target color chip, removed it from sight and, after 5 seconds, showed participants two color chips, one that was the same as the target, and one that was different. Participants were then asked to point to the same chip they had seen before. What they found was that participants were more accurate when the colors of the two chips belonged to different categories in their language than when the colors belonged to the same category (even if they were equally distant in both cases). And whereas English speakers were more accurate at recognizing color chips when the alternatives crossed the blue/green region, that effect was not observed in Himba or Berinmo speakers. Berinmo and Himba speakers performed better when the two color chips belonged to different categories in their own language, and blue and green are not different categories for them.

These findings go in the opposite direction of Berlin and Kay, and Heider, who claimed that there are certain universal colors that are conceived as focal regardless of the language one speaks. If Himba and Berinmo categorized blue and green as different colors, despite their language lacking that distinction, we would expect them to have a similar performance in the experiment to that of English

speakers. But that wasn't what the researchers observed. Himba and Berinmo speakers were better at memorizing colors that were focal in their own language, but not colors that were focal for English speakers.

But when it comes to the relation between color terms and perception, how should these findings be interpreted? Roberson et al., believe that their results show the "influence of language on perception. At the very least, our results would indicate that cultural and linguistic training can affect low-level perception (...). However, more than that, the results uphold the view that the structure of linguistic categories distorts perception by stretching perceptual distances at category boundaries." (2000, p. 394). So they go one step further than Kay and Kempton, for they believe that linguistic categories can alter color perception itself, that is, the way colors appear to us, and not just affect the *judgments* that we make about colors. But what is important for my purposes is that Roberson et al. agree with Kay and Kempton that *linguistic* categories are causally responsible for the differences observed (whatever exactly they are) in speakers of different languages.

Let's consider two more studies that found signs of the influence of language on perceptual tasks. One involved a comparison between Russian and English speakers. The Russian language has one word for lighter blues ("goluboy") and a different one for darker blues ("siniy"). That is, Russian, unlike English, forces their speakers to specify whether one is talking about dark blue or light blue. In a study by Winawer et al. (2007), Russian and English speakers were shown three color squares, and had to indicate, as quickly and accurately as they could, which of the two squares at the bottom was identical to the square on top. In order to test whether language had an influence on color discrimination, subjects performed the test under three conditions: one with no interference, one in which they had to simultaneously perform a verbal task, and one where they had to simultaneously perform a spatial task. What the researchers found was that Russian speakers showed a category advantage (being faster at discriminating colors when they belonged to different categories in their language than when they belonged to the same category) in trials without interference, and with spatial interference, but not with verbal interference. English speakers showed no categorical effects in either of the three conditions.

Winawer et al. are a little bit more cautious than Roberson et al. when it comes to interpreting how language acts on perception. They believe, similarly to Kay and Kempton, that linguistic representations somehow interfere with perceptual

decisions, given that the category advantage Russian speakers exhibited in normal trials was eliminated in trials with a verbal interference. According to them, the category advantage disappeared in trials with verbal interference because linguistic representations, that would normally be acting to help perceptual judgments, were being recruited for something else. As they put it,

our results suggest that language-specific distortions in perceptual performance arise as a function of the interaction of lower-level perceptual processing and higher-level knowledge systems (e.g., language) online, in the process of arriving at perceptual decisions. (...) it appears that language-specific categorical representations play an online role in simple perceptual tasks that one would tend to think of as being primarily sensory. (p. 7784)

The idea, then, is that judgments about colors (such as that *this* color is the same as *that* one) take into consideration more than just sensory information. Instead, linguistic representations, normally thought to be completely unrelated to perception, “normally meddle in even surprisingly simple objective perceptual decisions.” (p. 7784). So the authors don't go as far as Roberson et al., in saying that language distorts color perception proper – the experience one has when seeing colors. Linguistic categories may instead interfere in what they call “perceptual decisions”, in ways that resemble what Kay and Kempton called the “name strategy” – linguistic representations unconsciously act when we are faced with decisions about color identity. Despite the differences in their interpretation of the role of language in color perception, all the researchers seem to agree that linguistic categories are causally responsible for categorical perception.

It is worth considering one more study, which reports a different kind of evidence for the claim that language is actively involved in color perception. In a study by Gilbert et al. (2006), English speakers were instructed to fixate on a cross that was surrounded by twelve colored squares. The squares were all the same color except for one, the target, whose color differed slightly from the color of the other squares. Subjects had to indicate as quickly and accurately as they could on which side of the circle, left or right, the target was located. They alternated four

colors in the experiments (two shades of blue and two shades of green). In some trials, the target belonged to the same category as the other squares, being just a different shade of the same color (e.g. the target was blue and the other squares were a close shade of blue), while in other trials the target belonged to a different color category (e.g. the target was blue and the other squares were a close shade of green). They found that categorical perception only occurred when the target belonged to a different category from the other squares and was presented in the right visual field (RVF), but never when the target was presented in the left visual field (LVF). That is, the time it took participants to indicate that the target was on the left side of the circle was approximately the same, whether or not the category to which the target belonged was the same or different from the color category of the other squares – there was no categorical perception effect in the LVF. But when the target was presented on the right side of the circle, participants were faster to identify it when it belonged to a different category than when it belonged to the same color category as the other squares.

Given that the brain processes information from each visual field in the opposite cerebral hemisphere, and that the left hemisphere is the one thought to be predominantly involved in language processing, they explain the lateralization of categorical perception by saying that targets that appear in the RVF are processed faster by the left cerebral hemisphere, where linguistic representations can interfere. This means, according to them, that linguistic representations are being accessed and are interfering in discrimination when a target is presented to the RVF, but not when it is presented to the LVF. They also found that the categorical perception on the RVF disappeared when subjects had to perform the task with a verbal interference. They claim that their results “are consistent with the hypothesis that linguistic categories selectively influence color discrimination in the RVF. Color names modulated color discrimination, enhancing between-category distinctions and perhaps reducing within-category distinctions, but only when the target appeared in the RVF. These effects were disrupted by verbal interference.” (Gilbert et al., p. 490). They explicitly remain neutral as to whether “language alters the nature of the early visual response to colors” or whether it just “facilitates postperceptual processes” (p. 493). But like the other researchers, here again we see that they believe linguistic categories, or color names, are responsible for categorical perception of color.

All these studies suggest, against some universalist views, that we don't

all conceive and discriminate colors in the same way. The language we speak may interfere in the way we discriminate colors. But still, I believe there is room for an alternative explanation for what is causing the differences that have been observed across speakers of different languages (and in the same person depending on the visual field).

4. An alternative explanation

We have seen that different studies found evidence of categorical effects both in color memory and in color perception. Researchers don't seem to agree as to how exactly language interferes in color perception: some say that low-level color perception is altered by the linguistic categories of one's language, others say that language interferes in postperceptual processes, and others remain neutral. But they all agree that linguistic categories cause these categorical effects.

I believe, however, that it may be too rushed a conclusion to assume that color names (linguistic representations or categories) are interfering in the tasks involving color discrimination. The findings that the categorical perception of color has different patterns across speakers of different languages, that it can be disrupted by verbal interference, and that it only happens when a target is presented in the RVF are certainly interesting. But for all we know, what might be interfering with perceptual discriminations are color *concepts*, or categories, but not necessarily color *words* (or representations that are essentially linguistic). There are at least two reasons for thinking that this might be the case.

One reason to doubt that names, or linguistic representations, interfere with perceptual discrimination tasks is provided by Holmes and Wolff (2012). In their experiments, they observed categorical perception in the right visual field both with labeled and unlabeled (unnamed) categories of objects. They claim then that there is no evidence that lateralized categorical perception is driven by language, or by linguistic representations. According to them, the left hemisphere may be responsible for partitioning experience into categories, whether or not they have a name. As for the elimination of the categorical effect with verbal interference, they note that categorical perception "may be disrupted more by verbal than by spatial interference (e.g., Winawer et al., 2007) because the former disproportionately taxes not only linguistic processing but also left hemisphere processing

independent of language” (2012, p. 442). Their findings suggest, I believe, that color names are not necessary for the occurrence of categorical perception, for this can equally well happen when we discriminate between things whose categories don’t have a name. As they put it, “while it remains possible that CP might sometimes be driven by language, the more parsimonious conclusion is that CP, even for labeled categories, is driven by nonlinguistic factors.” (p. 442). Color concepts or categories, which are nonlinguistic, might then be what explains categorical perception.

Another reason for not assuming that linguistic representations are responsible for categorical perception effects is that some studies have found evidence for categorical organization of color in pre-linguistic infants as young as 4 months old (see Skelton et al., 2017). This is another type of evidence for the idea that the categorization of color doesn’t require language, given that even very young babies seem to group colors in categories. Instead of assuming that language, once learned, takes over and is responsible for categorical effects on perception and memory, it would again be simpler to assume that color concepts or categories are what explain the partitioning of colors in categories from the start.

So the conclusion that some researchers have extracted from their experiments, namely that linguistic representations interfere either directly with color perception or with postperceptual judgments about colors, seems unwarranted. What can be interfering in color perceptual discrimination and memory are color categories (or concepts), and not color words properly – for categorical perception may occur as a consequence of concepts that may or may not be named. Verbal interference could then disrupt categorical perception because it requires the use of other concepts. Likewise, categorical perception might occur in the right visual field and not the left because color concepts might be stored in the left hemisphere of the brain and can thus interfere in color discrimination more readily when they appear in the right visual field.⁶

⁶ Some studies have found activation in language regions of the brain when subjects performed color discrimination tasks (cf. Tan et al. (2008) and Siok et al. (2009)). However, as Holmes and Wolff (2012) note, this doesn’t show that color names are causally responsible for categorical perception effects. In one experiment by Tan et al. (2008) subjects had to indicate whether two colored squares were the same or different. In one condition, all colors were easy to name (good examples of red, green or blue), and in

In the case of tasks involving color memory, though, it may be more likely that color names (and perhaps not simply color concepts), are used by participants to improve their performance. Some tasks required subjects to look at a color chip, retain the color in memory, and then, after an interval, choose whether it was color chip A or chip B that they had seen. It is natural to assume that naming the color chip might be a strategy used to improve performance – for it might be easier to remember a color word than a color patch. As Pinker suggests, in experiments involving color memory, it could be that “subjects remembered the chips in two forms, a nonverbal visual image and a verbal label, presumably because two kinds of memory, each one fallible, are better than one.” (1994, p. 65). The retention in memory of color names could then be what explains the better performance when color chips A and B get different names. But when it comes to perceptual discrimination tasks that don’t involve memory, color concepts, and not words, might be doing the work.

But there is still something left to explain. Categorical perception effects are different across speakers of different languages, and consistent with the color names of their languages. What could explain that, if not color words? My suggestion is that, even if we assume that color concepts and not words are directly responsible for categorical perception effects, language is still part of the explanation for this phenomenon. We might assume that the process of learning color words may end up reshaping the extensions of our pre-linguistic color concepts, making them more or less inclusive than what they originally were. In learning the color words of our language we are being trained to make certain distinctions among colors and not others, that is, to pay more attention to certain differences than to others. This in turn might end up influencing our color concepts, or the way we categorize colors. People who speak a language with 5 color terms might then end up with different color concepts from people who speak a language

another they were hard to name. Brain regions associated with lexical search were active in the easy-to-name condition, but not in the hard-to-name. But this activation made no difference in their performance, for participants were equally fast and accurate in both conditions. Assuming color words were indeed being accessed in the easy-to-name condition, they had no power of making perceptual decisions faster or more accurate than in the hard-to-name condition, in which language areas were not active. Their activation might then be automatic, but unrelated to color discrimination.

with 11 color terms, because they typically distinguish certain colors and not others.

Unlike what most researchers have suggested, we don't need to assume that natural language words (or linguistic representations) are acting on color perception in order to account for the categorical effects that have so far been observed. Instead of assuming that color names are modulating either color perception directly, or perceptual judgments, an alternative is to assume that color concepts interfere with performance in color discrimination. But this doesn't require us to deny that language is part of the explanation of the variation we find in categorical perception effects, because the language we learn partly explains the color concepts we have. Words can have an important role in causally determining the extensions of our color concepts or categories, but they can still be conceived as dissociated from them – and as not determining performance in color perceptual discrimination tasks. That is, I think, a more plausible (or at a minimum an equally plausible) alternative interpretation of the findings.

One could here argue that we shouldn't stop the causal chain that explains our color concepts having the extensions that they have at color words. One could demand an explanation for the introduction of color words in a language, for why some languages have 2 color words whereas others have 11, and so on. Were we to trace back the origin of color words, we would get back to color concepts. People initially conceived colors in a certain way, or they for some reason found important to distinguish certain color categories, and that led to the appearance or borrowing of color terms. Pinker (2007), for instance, puts a lot of emphasis on this point, in arguing against strong versions of linguistic relativism, according to which language determines or shapes in a strong sense the way people think.

I believe, though, that my considerations are compatible with these observations. My emphasis has been on the causal role that color words play in the determination of the extension of our color concepts, from the perspective of an *individual* learning a language. Assuming that speakers of Russian and English end up having different color concepts in the blue region of the spectrum, a natural explanation for this difference is that it is due to the language they grew up speaking. But when we consider the issue of the relation between language and thought from a *historical* perspective, then I agree with Pinker that concepts are what explain the origin of color words. The two views are then compatible, because they highlight different aspects of the relation between words and concepts.

The view I'm suggesting then is not to be confused with a strong form of

linguistic relativism, according to which language is a necessary means for creating conceptual categories. On the contrary, I take language to be an important means to form categories, or to reshape preexisting ones, but by no means the only one. I assume that categorization can also occur and be altered by non-linguistic means.

Speakers of different languages habitually make different distinctions in order to speak about colors, which in turn might affect their color concepts or categories. But what is the nature of what I have been calling “color concepts or categories”? Are they perceptual? Or are they cognitive (or amodal)? I wish to remain neutral about whether these color concepts are categories that belong to perception, or whether they are the same concepts that are accessible to thoughts and reasoning about colors. I’m afraid I don’t have an answer to this question, simply because this is an empirical question, to be investigated by cognitive psychologists. The answer to this question bears on the topic of the encapsulation of color perception, but this discussion goes beyond the scope of this paper.

Finally, one thing that is important to highlight is that these experiments might not represent what happens in everyday life when we perceive, discriminate or remember colors. Participants had to solve tasks that involved noticing, remembering, or assessing differences or similarities between colors. Even if I’m wrong and color words, and not just color concepts, were in fact interfering in their performances in these tasks, that still doesn’t show that color words are recruited whenever we open our eyes and see and discriminate colors. It might well be that linguistic labels are recruited in the specific context of psychological experiments, in order to maximize performance. That it, the nature of the tasks participants had to perform might have favored the involuntary involvement of linguistic knowledge. And the same holds in fact if color concepts, and not color words, were causing the categorical effects. It might well be that these concepts were only used because of what the tasks demanded from participants, and not that they are regularly used in color discrimination as a default strategy. So it is unclear whether we can, from the experiments mentioned, generalize and claim that either color words or concepts are involved in everyday color discrimination and memory.

References

Allen, G. (1878). “Development of the Sense of Colour”. In: *Mind*, Vol. 3, No. 9, pp. 129-132

- Deutscher, G. (2010). *Through the Language Glass: Why the World Looks Different in other Languages*. New York: Metropolitan Books.
- Fodor, J. A. (1981). "The present status of the innateness controversy". In: *Representations: Philosophical Essays on the Foundations of Cognitive Science*. Cambridge, MA: MIT Press Bradford Books.
- Fodor, J. A. (1998). *Concepts*. Oxford: Oxford University Press.
- Gilbert, A. et al. (2006). "Whorf hypothesis is supported in the right visual field but not the left". In: *Proceedings of the National Academy of Sciences*, 103 (2) 489-494.
- Gladstone, W. E. (1858). *Studies on Homer and the Homeric age*. Vol. 3. Oxford: Oxford University Press.
- Heider, E. R. (1972) "Universals in color naming and memory". *Journal of Experimental Psychology*, Vol. 93, No. 1, 10-20.
- Holmes, K. J.; Wolff, P. (2012). "Does Categorical Perception in the Left Hemisphere Depend on Language?". *Journal of Experimental Psychology: General*, Vol. 141, No. 3, 439-443.
- Kay, P.; Kempton, W. (1984). "What is the Sapir-Whorf Hypothesis?" *American Anthropologist* 86: 65-79.
- Pinker, S. (1994). *The Language Instinct*. London: Penguin Books.
- Pinker, S. (2007). *The Stuff of Thought*. New York: Penguin Books.
- Roberson, D. et al. (2000). "Color Categories Are Not Universal: Replications and New Evidence From a Stone-Age Culture". In: *Journal of Experimental Psychology*, Vol. 129, No. 3, 369-398.
- Roberson, D. et al. (2005). "Color categories: Evidence for the cultural relativity hypothesis." *Cognitive Psychology*, 50, 378-411.
- Roberson, D.; Hanley, R. (2010). "Relatively speaking: an account of the relationship between language and thought in the color domain", in *Words and the mind: how words capture human experience*, B. C. Malt and P. Wolff (eds.). Oxford: Oxford University Press.
- Siok, W. T. et al. (2009) "Language regions of brain are operative in color perception", *Proceedings of the National Academy of Sciences* May 2009, 106 (20) 8140-8145.
- Skelton, A. E. et al. (2017). "Biological origins of color categorization". *PNAS*. 114 (21) 5545-5550.
- Tan, L. H. et al. (2008) "Language affects patterns of brain activation associated with perceptual decision". *Proceedings of the National Academy of Sciences* Mar

COLOR TERMS AND PERCEPTION

2008, 105 (10) 4004-4009.

Winawer, J. et al. (2007). "Russian blues reveal effects of language on color discrimination." In: Proc Natl Acad Sci. 2007; 104: 7780–7785.

WITTGENSTEIN, FORMALISM AND SYMBOLIC MATHEMATICS

ANDERSON LUIS NAKANO
USP

1. Introduction

In his essay '*On the Origin of Symbolic Mathematics and Its Significance for Wittgenstein's Thought*', Sören Stenlund tries to align Wittgenstein's approach to the foundations and nature of mathematics with the tradition of symbolic mathematics, which dates back to the works of Franciscus Vieta and René Descartes. According to Stenlund, "the symbolic view of mathematics offers us a perspective from which the unity of Wittgenstein's philosophy of mathematics becomes apparent" (2015: 25). This unity refers not only to a systematic coherence of Wittgenstein's remarks about mathematics, but also to a genetic continuity of his thought. Stenlund holds that Wittgenstein already defended a symbolic conception of mathematics in the *Tractatus*, and continues to defend such a view in his middle and late period as well.

The symbolic view of mathematics is contrasted in Stenlund's essay with the Greek or the so-called "ontological conception" of mathematics¹. This contrast is developed in three interrelated levels: the level of mathematical objects (numbers, in particular), the level of mathematical systems (e.g., arithmetic, algebra) and, finally, the level of the relation between mathematics and ordinary verbal language. In each of these levels the symbolic approach to mathematics is meant to be illuminated by means of a comparison with the ontological approach.

The conception of number is characterized in the ontological approach by its essential relation to a determinate multitude of things. Central to this conception is the idea that numbers do not exist *per se* as elements of a calculus, but are

¹ The terminology is due to Klein (1968).

always numbers *of* a given plurality of things. On the other side, the symbolic view holds (according to Stenlund) that numbers are nothing but symbols within a calculus, determined by how we operate with numerals in arithmetic. Because numerals are identified as pieces inside a calculus, arithmetic is not considered to be a theory or doctrine *about* something, but as an activity, a “pure calculus”. And this holds true in the symbolic view for all mathematical systems, including logic and geometry. This implies that, on one side, logic does not hold any special position among the calculi, and cannot be a theory *about* “correct reasoning” or “correct inferring”. On the other side, geometry is not a theory about space (physical or phenomenological), but only a calculus side by side with others. Another consequence of this view is that the notion of truth in mathematics does not gain its significance through the relations that a mathematical system could possibly maintain with what it is about. The notion of truth is, therefore, internal to the system, and it is related to the operational practices that constitute it.

Regarding the relation of mathematics and ordinary verbal language, the ontological approach present in Greek mathematics is characterized by a continuity and proximity with language in its verbal form. Stenlund argues that this view is still present in foundational writers, which have the tendency “to give meaning and significance to basic notions in mathematics and formal logic by translation or paraphrase into verbal language” (2015: 37). And this tendency is, for Stenlund, responsible for the phenomenon that Wittgenstein labels “prose” (*Prosa*), the “everyday prose that accompanies the calculus” (WWK: 129), which is regarded by Wittgenstein as the main source of the confusions about the foundations of mathematics. Stenlund presents the symbolic conception of mathematics in sharp contrast to this tendency. According to him, the symbolic conception holds that mathematical symbols have content only because (and just to the extent that) they are part of a mathematical system, of a calculus, and not because they are related to and continuous with ordinary language. Mathematics is, therefore, autonomous and independent of everything that is external to the calculus, including its possible external applications. In Stenlund words: “an essential feature of the symbolic point of view was the *logical separation* of a symbolic system from its application to some subject-matter outside pure mathematics” (2015: 46, my emphasis).

This characterization of symbolic mathematics raises naturally the question of whether the symbolic view coincides with formalism, in the sense of Thomae (and not in the sense of what is called “Hilbert’s formalism”). Stenlund

argues that the problem with this identification is the superficial and pejorative sense that the word “formalism” is used by authors like Frege and Brouwer. If, however, we disregard this negative attitude towards the word “formalism” and the misconstruals of this conception², then the identity between symbolic mathematics and formalism is, according to Stenlund, appropriate. He goes on to say that that Thomae's concept of “formal arithmetic” is “one of the most clear and distinct examples of the use of the symbolic point of view” (2015: 49).

Now, if the symbolic conception of mathematics is identical with formalism and is, at the same time, the perspective from which we can consider the unity of Wittgenstein's remarks on mathematics, the question to be asked is of course whether Wittgenstein holds after all a formalist position in philosophy of mathematics. The present paper addresses this question, with a particular emphasis on Wittgenstein's views in his middle period. Although the discussion of the relations between Wittgenstein's views and formalism is certainly not a novelty, I think that there is no presentation in the literature that adequately explains Wittgenstein's trains of thought in his middle period about the applicability of arithmetic (this point being the crux of the debate between Frege and the formalists). Another point worth emphasizing is that I shall not argue here with Stenlund's characterization of symbolic mathematics, but only take, as point of departure, his reading according to which “formal”, as used by Thomae, is *essentially synonymous* with “symbolic”, as this word has been used in the tradition of symbolic mathematics³. However, the way I will try to show why Wittgenstein should not be read as a formalist *could be useful*, I think, for a better appraisal of

² Frege, for instance, makes an effort to put Thomae's views together with Heine's views, according to which arithmetic is about the signs themselves. This is clearly a misconstrual of Thomae's views, and this misconstrual is the responsible for the dichotomy Frege sustained between arithmetic conceived as referring to abstract objects and arithmetic conceived as referring to the signs used in arithmetical calculations. See, in particular, §88 and §§95-6 of Frege (1903). Shapiro (2000) distinguishes Thomae's views from Heine's by calling the first a “game formalism” and the second a “term formalism”.

³ See Stenlund (2015: 49): “One of the most clear and distinct examples of the use of the symbolic point of view is the mathematician Johannes Thomae's concept of “formal arithmetic” (...). As Thomae uses the word ‘formal’ it is essentially synonymous with ‘symbolic’, as this word has been used in mathematics ever since Vieta”.

the relations between Wittgenstein's views on mathematics and symbolic mathematics. In his characterization of symbolic mathematics, Stenlund focuses too much, I think, on the *operational* aspects of the symbolism, and very little on the *esthetic* aspects of it and its relations with the notions of scheme, paradigm and aspect. These notions will play an important role in my argument against the characterization of Wittgenstein as a formalist.

In the next section, I will start with some remarks about the approximation of Wittgenstein's philosophy of mathematics and formalism, indicating why the reading of Wittgenstein as a formalist is more inviting when we consider the writings from his middle period. Section 3 tries then to remove the temptation to regard the middle Wittgenstein as a formalist. Section 4 casts light on the reasons Wittgenstein had, in his middle period, to think that arithmetic guarantees alone its applicability, and shows that Wittgenstein's idea of the autonomy of arithmetic is not the formalist idea of autonomy according to which a calculus is logically separated from its applications. Section 5 turns to the case of algebra, which is contrasted with arithmetic in that, unlike arithmetic, it is not an autonomous mathematical system. Section 6 is devoted to some concluding remarks.

2. Wittgenstein and formalism

Stenlund was not the first to characterize Wittgenstein's attitude towards mathematics as closer to formalism. Rodych (1997) is even more explicit and suggests that Wittgenstein is to be read as advocating a formalist stance. Rodych, however, points out that there are some subtleties that need to be considered to properly characterize the development of Wittgenstein's views about mathematics. For this reason, he draws a distinction between two varieties of formalism: strong formalism and weak formalism⁴. They are defined as follows:

Strong Formalism (SF): A mathematical calculus is defined by its accepted or stipulated propositions (e.g., axioms) and rules

⁴ Rodych (1997) also defines other two varieties of formalism, namely, extreme formalism and radical formalism, but these variants of formalism are not attributed to Wittgenstein.

of operation. Mathematics is syntactical, not semantical: the meaningfulness of propositions within a calculus is an entirely intrasystemic matter. A mathematical calculus may be invented as an uninterpreted formalism, or it may result from the axiomatization of a “meaningful language.” If, however, a mathematical calculus has a semantic interpretation or an extrasystemic application, it is inessential, for a calculus is essentially a “sign-game” – its signs and propositions do not refer to or designate extramathematical objects or truths.

Weak Formalism (WF): A mathematical calculus is a formal calculus in the sense of SF, but a formal calculus is a mathematical calculus only if it has been given an extrasystemic application to a real world domain. (Rodych 1997: 196-7)

Rodych argues that both in the *Tractatus* and in RFM Wittgenstein adopted weak formalism, whereas in his middle period Wittgenstein espoused strong formalism. In this essay, I will focus on criticizing the view that, in this middle period, Wittgenstein was a formalist (strong or weak), because it is in this period that a formalist reading of his writings is more tempting. With regard to the late period, even if Rodych is right on characterizing Wittgenstein's position as a weak formalism⁵, from this it would follow, contrary to what Stenlund says, that there is no “logical separation” of a mathematical system from its application to some subject-matter outside pure mathematics, because it is only by being “logically connected” with (at least one of) its applications that a formal system deserves to be called mathematics. In this connection, Rodych quotes the well known passage of RFM in which Wittgenstein states that “it is the use outside mathematics, and so the *meaning* of the signs, that makes the sign-game into mathematics” (RFM V, §2). Now, Stenlund also recognizes that, for the late Wittgenstein, “the use of mathematical signs in applications outside mathematics contributes to the meaning of mathematical signs” (2015: 63-4). But then it is unclear what is the “logical separation” of a symbolic system from its application to some subject-matter outside pure mathematics. We should then conclude that Wittgenstein was not in agreement with this “essential

⁵ But see Mühlholzer (2010: 72ff) for some misgivings about this characterization.

feature of the symbolic point of view” according to Stenlund and, therefore, that this view is not a good perspective to consider the unity of Wittgenstein's thought about mathematics.

A second misgiving one might well have about calling Wittgenstein a formalist arises from his characterization of mathematical propositions as *grammatical*. In his essay, Stenlund argues as if the grammatical view of mathematical propositions reinforced Wittgenstein's “symbolic, non-ontological conception of mathematics” (2015: 25). But in which sense “grammar” can be said to be logically separated from its applications?⁶ If we conceive of mathematical propositions as rules of grammar, it is just not the case that pure mathematics is non-descriptive and applied mathematics is descriptive. As Düring puts it: “*c'est en tant qu'elles s'appliquent que les mathématiques ne traitent de rien*” (Düring, 2005: 203).

The view that mathematics does not need to be descriptive to guarantee its applicability (*Anwendbarkeit*) was already Wittgenstein's view in the *Tractatus*. Now, it is controversial whether the *Tractarian* view of mathematics could be called “formalist” even in the weak sense given by Rodych. For it is not that there is a formal system, namely, arithmetic, that becomes mathematics when one shows that it has an extrasystemic application. What is given by the connection between number and operation in section 6.02 (where number is defined as the exponent of an operation) is not an external application of a “pure calculus”, but the *form* of such an application. Similarly, what is given by the definition of the general form of an operation in section 6.01 is the general form of all *possible* applications of numbers. Thus in the *Tractatus* the applicability of a calculus is guaranteed and its limits

⁶ As far as I can see, Stenlund's chain of reasoning throughout his essay is roughly the following: because mathematics is symbolic, and therefore non-ontological (not about something), it is cut off from its extra-mathematical applications. This reasoning in turn could be regarded as a kind of *modus tollens* of Frege's view about mathematics. For Frege, because mathematics has to guarantee all its applications (whether internal or external), mathematics has to be about something. Now, it seems to me that the conception of mathematical propositions as grammatical is an attempt to resist this implication, and not an attempt to deny the conclusion by means of a contrapositive to Frege's reasoning. Wittgenstein may be read, thus, as pointing out that it does not follow from the non-descriptive character of mathematics that mathematics is cut off from its applications.

established even though we still do not have a single example of an extrasystemic application to a “real world domain”. But there is surely no “logical separation” of arithmetic from its external applications as well.

What about the writings from Wittgenstein's middle period? I must confess that, in these writings, there are some passages that strongly suggest Rodych's interpretation according to which Wittgenstein defended a strong formalism about mathematical calculi. In PR, for instance, Wittgenstein wrote:

You could say: Why bother to limit the application of arithmetic, that takes care of itself. (I can make a knife without bothering which sorts of material it will cut: that will show soon enough.) (PR X, §109j)

It's always a question of whether and how it's possible to represent the most general form of the application of arithmetic. And here the strange thing is that in a certain sense it doesn't seem to be needed. And if in fact it isn't needed then it's also impossible. (PR X, §110b)

Because Wittgenstein is here clearly giving up the *Tractarian* account of the applicability of arithmetic, saying on the one hand, that it is superfluous and, on the other hand, that it is impossible, this last passage seems to suggest that Wittgenstein is cutting the pure arithmetical calculus off from its applications. Likewise, when he says that the application takes care of itself, he seems to be saying that we cannot guarantee *a priori* that a mathematical calculus could be applied, but at most say that it “can be applied only to what it can be applied to” (WWK: 104). The applicability of a formal calculus of mathematics would then be a *contingent* matter, and not something that could be guaranteed beforehand.

It is, thus, tempting to read the middle Wittgenstein as a strong formalist. In what follows, I shall argue that this temptation should be resisted, however. While it is true that, in his middle period, Wittgenstein came to realize that the *Tractarian* account of the applicability of arithmetic is not needed, I shall argue that this change in his thought does not imply that he started to regard the applicability of arithmetic as a contingent or external matter. As we shall see in the next section, Wittgenstein still sticks in his middle period to the idea that the applicability is an essential component of mathematics.

3. The applicability as a criterion for mathematics proper

Let me return for a moment to Stenlund's characterization of symbolic mathematics and its relation to formalism. One page after having raised the question whether the symbolic view of mathematics coincides with formalism, Stenlund quotes a nice passage from Couturat and qualifies this passage as a "very pertinent formulation of the symbolic view of mathematics" (2015: 52). The passage in question is the following:

A mathematician never defines magnitudes [or numbers] in themselves, as a philosopher would be tempted to do; he defines their equality, their sum and their product, and these definitions determine, or rather constitute, all the mathematical properties of magnitudes. In a yet more abstract and more formal manner he lays down symbols and at the same time prescribes the rules according to which they must be combined; these rules suffice to characterize these symbols and to give them a mathematical value. Briefly, he creates mathematical entities by means of arbitrary conventions, in the same way that the several chessmen are defined by the conventions which govern their moves and the relations between them. (Quoted in Bell 1937: 624)⁷

This characterization of the symbolic view of mathematics somewhat coincides with strong formalism as defined by Rodych. Under this picture, mathematical signs are, like chess pieces, elements of a calculus-game and they receive their content as symbols by means of the rules that define this game. In this sense, it is enough to determine the operations with mathematical signs to fill these signs with mathematical content and give them mathematical citizenship. In arithmetic, for example, there is no need to say what numbers are, but only how they are moved

⁷ The addition between brackets is from Stenlund.

in the arithmetic-game.

Now, if the middle Wittgenstein is read as a strong formalist, then it is difficult to understand some passages in which he is apparently not satisfied with the fact that some rules for the manipulation of signs are laid down, that he thinks this is not enough to consider a symbol or an operation as *mathematical*. Commenting on intuitionistic choice sequences, for instance, he states the following:

If Weyl believes that [a freely developing sequence] is a mathematical structure because I can derive a freely developing sequence from another by means of a general law, e.g.,

$m_1, m_2, m_3, \dots$

$m_1, m_1 + m_2, m_1 + m_3, \dots$

then the following is to be said against this: No, this shows only that I can add numbers, but not that a freely developing sequence is an admissible mathematical concept. (WWK: 83)

The last sentence would be a very curious remark from the perspective of a formalist. Indeed, if we are given laws for manipulating choice sequences, for computing their sum, their product, and so on, what else is required from a formalist standpoint? Another example is Wittgenstein resistance to accept the so-called pseudo-irrationals (such like π' , P and F^8) as *bona fide* arithmetical symbols. It is unclear why a formalist should have any qualms with these perfectly computable objects.

A further difficulty to read the middle Wittgenstein as a formalist is that,

⁸ In BT, §139, Wittgenstein defines these pseudo-irrationals as following: “ π' is a rule for the formation of decimal fractions; specifically, the expansion of π' is the same as the expansion of π except where the sequence 777 occurs in the expansion of π ; in that case the sequence 000 replaces the sequence 777. There is no method known to our calculus for discovering where we will encounter such a sequence in the expansion of π . P is a rule for the construction of binary fractions. At the n th place of the expansion there occurs a 1 or a 0, depending on whether n is prime or not. F is a rule for the construction of binary fractions. At the n th place there is a 0, except when a triple x, y, z from the first 100 cardinal numbers satisfies the equation $x^n + y^n = z^n$.”

contrary to what is expected from a formalist, he seems to answer what a number is, i.e., he seems to succumb to the temptation warned by Couturat in the passage above. In PR we are told that “numbers are pictures (Bild) of the extensions of concepts” (PR X, §100a), that “a cardinal number is an internal property of a list” (PR XI, §118d) and that “a real number is: an arithmetical law which endlessly yields the places of a decimal fraction” (PR, §186b). Why should a formalist bother himself with defining what numbers are? Are not numbers just what is determined by how we manipulate numerals?

Still, if the middle Wittgenstein is not a formalist about mathematics, what does he think that is required for mathematics besides a system of rules for manipulating signs? I think that here Wittgenstein would repeat Frege almost verbatim: it is the applicability alone which elevates a system of rules from a game to the rank of mathematics (see Frege, 1903, §91)⁹. Contrary to Frege, however, Wittgenstein does not require that mathematical sentences express thoughts in order to be applied. As I shall argue, in his middle period Wittgenstein thought that arithmetic could guarantee its applicability completely a priori by being its own application. In this sense, the lemma “the application of arithmetic takes care of itself” is not to be read as saying that the applicability of arithmetic is a contingent matter, but as saying that arithmetic is autonomous (autonom), because it guarantees by itself its own applicability (thus providing its own *raison d'être*).

At this point, it may seem that I am being unfair to Rodych's characterization of Wittgenstein as a formalist. This is because Rodych does not deny that, for a formalist, a mathematical system could have internal or intrasystemic interpretations. He also emphasized Wittgenstein dictum that “arithmetic is its own application” (see Rodych, 1997: 199). However, Rodych interprets this dictum in such a way that it becomes valid for all system of rules for manipulating signs whatsoever. The idea of an internal or intrasystemic application

⁹ See Medina (2002: 87): “Despite the influence of formalist ideas on Wittgenstein's view, the primary concern of his discussion of rules is the issue of applicability emphasized by Frege. In this respect Wittgenstein's view of rules is more indebted to Frege's critique of formalism than to formalism itself”. Medina, however, does not provide, in my opinion, a satisfactory reading of the passages where Wittgenstein says that arithmetical constructions themselves guarantee their applicability. His reading, for instance, is unable to explain Wittgenstein's qualms about the applicability of a pseudo-operation like $\times'$ (see p. 9 below).

is, for him, “the strong formalist idea that the meaningfulness and the truth or falsity of propositions within a calculus are determined entirely by the axioms and rules of operations of that calculus, without any necessary reference to an extrasystemic application” (1997: 200). In this sense, the requirement of applicability of a calculus becomes tautologous, because it is satisfied by any calculus. I shall argue, on the contrary, that the idea that arithmetic is its own application is a substantial feature of arithmetic and, moreover, that this feature is important to understand Wittgenstein's thought about mathematics in his middle period.

There is a passage in PR where the non-tautologous character of the applicability of a calculus becomes apparent, namely: “It is clear that were I able to apply $\times$ ’ all doubts about its legitimacy would be dispelled. For the possibility of application is the real criterion for arithmetical reality”¹⁰ (PR XVII, §186f). Now, it is obvious that I can apply the rule $\times$ ’ in Rodych's sense. For the formalist, the operation $\times$ ’ should be an operation for manipulating signs in equal footing with any other operation. If Wittgenstein raises doubt about the applicability of this rule, then his notion of application is not the same as the application of a rule inside a calculus-game. In the same passage, Wittgenstein says, I repeat, that it is the applicability the real criterion for arithmetical reality. I would say that this criterion holds, in Wittgenstein's middle period views, not only for arithmetic, but for mathematics in general. This point will become particularly evident when we later consider the case of algebra, because algebra, unlike arithmetic, is not “its own application”, depending for its “mathematical existence” on the donation of sense (applicability) for its formulae. Arithmetic, by contrast, guarantees its own existence by being its own application. Arithmetic is, in this sense, autonomous, not because it is cut off from its external applications, but because it is *causa sui*. The next section is devoted to explain this idea of autonomy of arithmetic.

4. The autonomy of arithmetic

The idea that arithmetic is autonomous is explored in PR together with the idea that arithmetic is a “kind of geometry”. Both in the case of arithmetic and in the case of

¹⁰ By “ $\times$ ” Wittgenstein means the operation which is the same as ordinary multiplication except for the fact that, in the result, every occurrence of the digit ‘7’ is replaced by ‘3’.

geometry it is possible to say, according to Wittgenstein, that they are their own application. This shows that they alone guarantee their own applicability:

One always has an aversion to giving arithmetic a foundation by saying something about its application. It appears firmly enough grounded in itself. And that of course derives from the fact that arithmetic is its own application. (PR X, §109a)

You could say arithmetic is a kind of geometry; i.e. what in geometry are constructions on paper, in arithmetic are calculations (on paper). – You could say it is a more general kind of geometry. (PR X, §109h)

The point of the remark that arithmetic is a kind of geometry is simply that arithmetical constructions are autonomous like geometrical ones, and hence so to speak themselves guarantee their applicability. / For it must be possible to say of geometry, too, that it is its own application. (PR X, §111a)

The main difficulty of these passages is to explain in which precise sense arithmetic (and geometry) could be said to be its own application. The following passage from the conversations with the Vienna Circle recorded by Waismann, where Wittgenstein says that this idea is “tremendously important”, can be helpful:

Mathematics is its own application. This is tremendously important. A lot follows from it. When I say '3 plums + 4 plums = 7 plums', '3 men + 4 men = 7 men', etc., I do not apply numbers to different objects; it is always the same application that I have before my eyes. Numbers are not represented by proxies; numbers are there. Only objects are represented by proxies. / The correctness of an arithmetical proposition is never expressed by a proposition's being a tautology. In the Russellian way of expressing it, the proposition $3 + 4 = 7$ for example can be represented in the following manner:

$(E3x)\phi x . (E4x)\psi x . \sim (\exists x)\phi x . \psi x : \supset: (E7x).\phi x \vee \psi x.$

Now one might think that the proof of this equation consisted in this: that the proposition written down was a tautology. But in order to be able to write down this proposition, I have to know that $3 + 4 = 7$. The whole tautology is an application and not a proof of arithmetic. Arithmetic is used in constructing this proposition. The fact that a tautology is the result is in itself inessential. For I can apply an arithmetical equation both to propositions with sense and to tautologies. (WWK: 35)

In the passage above, Wittgenstein is saying that from the fact that mathematics is its own application, two extremely important things follow, namely:

1. all applications of arithmetical equations are, in a sense, the same application (this being connected with the non-surrogative character of numbers).
2. the tautology correspondent in Russell's symbolism is not the proof of the equation, but an application of the equation. As a consequence, the tautology is not needed to apply the equation, since it is already one of its applications.

It is important to stress here the direction of the reasoning. It is not because the tautology in Russellian symbolism is not needed for the application of arithmetic that arithmetic is its own application, but quite the opposite: it is because arithmetic is its own application that we do not need to bridge a supposed gap between arithmetic and its applications by means of a logical schema like the tautology above. The same direction is emphasized in the following passage of PR: "Every mathematical calculation is an application of itself and only as such does it have a sense. That is why it isn't necessary to speak about the general form of logical operation when giving a foundation to arithmetic" (PR X, §109e). Again: it is because arithmetic is its own application that the Tractarian explanation for the applicability of arithmetic is not needed.

I must point out at this moment that, when Wittgenstein says, in the passages above, that "mathematics is its own application" and that "every mathematical calculation is an application of itself", the term "mathematics" should

be read, according to my interpretation, contextually as “arithmetic”. This is because, according to the reading I will suggest to these passages, it is simply not true that every mathematical calculation is an application of itself. Algebraic calculations, for instance, are not their own application (I will return to this point later).

Another point worth noticing in this last quotation from PR is that the idea of mathematics being its own application is rephrased and interpreted as the self-applicability of mathematical calculations to themselves. This is a bit puzzling, since this notion of self-applicability, as far as I know, does not occur elsewhere in the writings from the middle period, except for a single passage of MS 107 where Wittgenstein says that “the calculation with strokes is also, at the same time, an application of the calculus. This ceases to be the case in this direct way in decimal system”¹¹ (MS 107: 68; Wi2: 40, my translation). Despite the scarce textual evidence, I think that this last passage provides an important hint to the direction where the answer to the question “what does it mean to say that arithmetic is its own application?” could be searched for. For it tells us that this idea of self-applicability is present in a “direct way” in calculations with strokes and that this immediacy is lost when we calculate in decimal system.

In a short text entitled “What is a number” written between 1929-30¹² (and intended to present Wittgenstein's views), Waismann presents this immediacy feature of the stroke-notation in connection with the idea of pictoriality (*Bildhaftigkeit*):

In speaking of 5 men I can represent the men by strokes.
But those men's being 5 is not represented; it manifests
itself in the fact that the number of strokes is 5. Here the
number sign is immediately conceived as a picture.
(WWK: 225, my emphasis)

This in turn explains how numbers are, as Wittgenstein says, “pictures of the extensions of concepts”: they are forms which present themselves both in the fact

¹¹ Original: “Das Rechnen mit Strichen ist zugleich auch eine Anwendung der Rechnung. Das hört in dieser direkten Weise im Dezimalsystem auf”.

¹² See WWK, p. 20-1.

(in the symbol) we use to represent the extension of a concept and in this extension itself. In stroke-notation, these forms are at the same time symbolized by means of the notation and exemplified in the notation. That is, in stroke-notation these forms are ostensively exhibited in the notation. This feature is lost (in this direct way) when we work in decimal notation. In the continuation of his text, Waismann makes clear, however, how and in which sense this feature is present even in decimal notation, albeit indirectly:

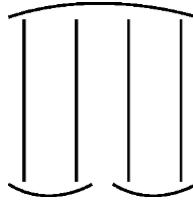
The usual way of representing the numbers by means of the system of digits rests on exactly the same principle. At first blush the number 387 does not seem to be a picture of the quantity it means. We must, however, take into account that in addition to the signs there are the rules of syntax too. The signs 3, 8, 7 are defined. If we reduce them to their definitions, that is, if we analyse these signs step by step, then they assume the very multiplicity they mean; e.g. $3 = 1+1+1$. Second, the position of figures, too, depicts something. Our number signs contain the possibility of being transformed into other signs that are pictures in an immediate way. That is, our number signs, together with the rules of syntax, are instructions for the construction of picture-like symbols.

There must always remain a clear way back to a picture-like representation of numbers leading through all arithmetical symbols, abbreviations, signs for operations, etc. The symbolism of the representation of numbers is a system of rules for translation into something picture-like (Bildhafte). (WWK: 225-6)

In short, numbers are forms and they occur in arithmetical symbolism either directly (in stroke-notation) or indirectly (in decimal notation plus the rules of syntax). These forms are not represented by proxies, but they “are there”. In each application of numbers, these forms are there in the same way that they are there in arithmetical symbolism itself. In this sense, the symbolism itself is an application of these forms (an instantiation of them).

Now, when we calculate with strokes in stroke-notation, the manipulation

of these strokes (grouping some of them together, separating others, etc.) can function as pictures of possible states of affairs. This is because the form of these possible states of affairs is represented as a structure in the symbolism itself. For instance, when I prove that $2 + 2 = 4$ in stroke-notation by means of the following diagram:



then this diagram could be used as a propositional sign for the projection of a situation in which 4 objects are grouped in 2 and 2, i.e., the diagram can *picture* such a situation. As in the *Tractatus* (cf. aphorism 2.201), the picture depicts (*das Bild bildet*) reality by representing a possibility of existence and non-existence of a determinate state of affairs, in this case, the grouping of four things in two and two. But for this to be possible both the picture and the state of affairs must share the *same* pictorial form (*Form der Abbildung*). This form is presented as an actual structure in the symbolism and as the possibility of this structure at the level of the things themselves¹³.

In Chapter X of PR, Wittgenstein insists on the fact that it is not necessary for a multitude of things to be presented as the actual extension of a concept. They could be presented by means of a list, say, $\{a, b, c, d\}$ (and the use of '=' to construct a concept, namely, " $\xi=a \vee \xi=b \vee \xi=c \vee \xi=d$ ", is only a subterfuge). That is why it is more generally valid to say that the cardinal number is an internal property of a list than to say that cardinal numbers are "pictures of the extensions of concepts". When an extension is given by means of a concept, then the cardinal number is such a picture, but when we have just a list of things, the cardinal

¹³ See PR (X, §100d). Note, however, that the "things" are not objects in the Tractarian sense, but ordinary things (plums, men, apples, and so on).

number is an internal property of this list¹⁴.

In stroke-notation, these internal properties are exemplified in the symbolism for numbers. If we say that a “scheme” or a “paradigm” is the “explicit symbolic exemplification of an internal property” (Narboux, 2001: 583) and if we say, moreover, that an “aspect” is the “internal relation exemplified within a scheme by the scheme itself” (*ibid*: 583), then we can say that, in stroke-notation, numbers are aspects of number-schemes or number-paradigms. That is, in stroke-notation numbers are symbolized by means of an application of them, namely, to strokes¹⁵. If I am right, it is this paradigmatic feature of arithmetic (presented directly in stroke-notation or indirectly in other notations) that is being targeted by Wittgenstein when he says that “arithmetic is its own application”. And this is, in turn, what makes arithmetic autonomous and guarantee *a priori* its applicability.

When we work in other notations (like the decimal system), the paradigmatic feature of arithmetical operations is not immediately present, but is present if we consider the calculations and the rules for manipulating symbols in these notations. But these calculations need to be always translatable in stroke-notation. The problem with a pseudo-operation like $\times'$ mentioned above is that there is nothing corresponding to it in stroke-notation. In this case, the decimal system ceases to be a mere mode of presentation of what is being considered (namely, arithmetical relations) and becomes itself an object of consideration. But then the whole symbolic chain that connects this notation with stroke-notation is lost and, as a result, it becomes doubtful what is the application of such symbolic manipulations.

5. The case of algebra

¹⁴ For Wittgenstein, a property is internal to an object just when “it is *unthinkable that its object should not possess it*” (*Tractatus*, 4.123).

¹⁵ Frascaola (1994: 44-54) also points out the relation between the arithmetic of strokes and the notions of paradigm and aspect. However, Frascaola (2004) does not draw the connection between these notions and Wittgenstein's remarks about the autonomy and self-applicability of arithmetic.

Wittgenstein discussed the relation between arithmetic and algebra¹⁶ for the first time in his middle period in the occasion of his detailed discussion of Skolem's recursive proof of the associative law: $a + (b + c) = (a + b) + c$ (abbreviated by Wittgenstein as "A(c)"). This discussion was undoubtedly regarded by Wittgenstein as extremely relevant to the philosophy of mathematics, and it would be exposed in many writings from his middle period, most importantly in Chapter XIV of PR and the whole Section of BT entitled "Inductive Proofs. Periodicity". I shall here consider just some passages of PR that are important for the reading I am suggesting of the autonomy of arithmetic. These passages show, I think, that Wittgenstein did not consider algebra as an autonomous calculus in the sense I have given to the notion of autonomy.

The upshot of the discussion of inductive proofs in PR is that induction is not to be conceived as the proof of an algebraic proposition, but as what shows the applicability of algebraic formulae to arithmetic¹⁷. In this vein, Wittgenstein warns us against confusing "the infinite possibility of its application with what is actually proved" (PR XIV, §163e). The so-called inductive or recursive "proof" only gives us the unlimited possibility of applying an algebraic formula to numbers, but it does not prove a general proposition about all numbers:

A recursive proof is only a general guide to an arbitrary special proof. A signpost that shows every proposition of a particular form a particular way home. It says to the proposition $2 + (3 + 4) = (2 + 3) + 4$: 'Go in this direction (run through this spiral), and you will arrive home.' (PR XIV, §164a)

To what extent, now, can we call such a guide to proofs the proof of a general proposition? (Isn't that like wanting to ask: 'To what extent can we call a signpost a route?') Yet it surely does justify the application of A(c) to numbers. (PR XIV, §164b)

¹⁶ By "algebra" Wittgenstein means the calculus with characters (*Buchstaben*), these characters being understood not as variables, but as constants.

¹⁷On this point, see also WWK, p. 33-4 and LWL, p. 18-9.

An induction doesn't prove the algebraic proposition, since only an equation can prove an equation. But it justifies the setting up of algebraic equations from the standpoint of their application to arithmetic. (PR XIV, §167d)

Notice that, in the last two passages above, Wittgenstein is saying the the application of the algebraic formula $A(c)$ to numbers is *justified* (*rechtfertigt*) by the induction. In this sense, the applicability of $A(c)$ does not “take care of itself”. We need to see the induction in order to see the applicability of $A(c)$ to numbers. In contrast to this, we do not need to see anything beyond the forms exhibited by the arithmetical symbolism itself in order to see the applicability of arithmetical equations to, say, plums or men.

In the sequence of the same chapter, Wittgenstein remarks that the algebraic formula only obtains its sense (its mathematical *existence*) by means of induction:

An algebraic schema derives its sense from the way in which it is applied. So this must always be behind it. But then so must the inductive proof, since that justifies the application. (PR XIV, §167a)

Through them alone the algebraic system becomes applicable to numbers. And so in a particular sense they are certainly the expression of something arithmetical, but as it were the expression of an arithmetical existence. (PR XIV, §167f)

If we remember that Wittgenstein equated mathematical existence (mathematical reality) with applicability, then it is clear from this last passage that Wittgenstein is saying that algebraic formulae, unlike arithmetical and geometrical, do not themselves guarantee their applicability, but need to be supplied with inductions.

In the same paragraph 167 of PR, Wittgenstein uses the pair of notions “sense/truth” to characterize the relation of induction to algebraic axioms. On one

side, algebraic equations are not proved by induction and therefore do not acquire their truth through it. On the other side, induction gives *sense* to algebraic equations: it makes possible the application of algebra to arithmetic. Another pair of notions used by Wittgenstein to explain this relation is the pair “name / proposition”. As stipulations, algebraic axioms like $A(c)$ are more like *names*: they are not bipolar like propositions (another reason to say that the induction does not prove it). But a name is only a name when it “denotes” something, that is, it receives its content only by pointing to its reference. Analogously, $A(c)$ only receives its content by pointing to an induction.

I conclude that, for the middle Wittgenstein, algebra is not “its own application” in the sense arithmetic and geometry are and, thus, algebra is not autonomous (its existence does not follow from its essence). That is why, in the case of algebra, we need to bridge the gap between its formulae and its applications. According to Wittgenstein, it is induction that bridges this gap between algebra and arithmetic.

6. Concluding remarks

In the preceding sections, I have intended to show why Wittgenstein should not be read as advocating a formalist view of mathematics. While he did use the formalist metaphor of a game with signs to dispel some misconceptions about the nature of mathematics, he still maintained (even in his middle period) Frege's idea that the applicability is something essential to mathematics proper. If I am right, Rodych's interpretation of the middle Wittgenstein as a (strong) formalist is misconceived. Arithmetic is autonomous not in the sense that it is essentially a sign-game with contingent applications, but in the sense that it displays its application in its own symbolism.

To return to Stenlund's characterization of Wittgenstein's conception of mathematics as “symbolic”, it must be emphasized that it does not follow from the fact that mathematics is not “ontological” (not a theory about an independent reality) that it is logically separated from its applications. The assumption behind this implication is that the axioms of a mathematical system become true (descriptive) propositions when applied to a “real world domain”, that we are able to find many applications of a mathematical system by finding true relations in some domain that

“match” those axioms. And then, if mathematics is not about something, it has to be cut off from its applications. But it seems to me that it is precisely this picture of the notion of application that Wittgenstein is trying to oppose:

If we say “it must be essential to mathematics that it can be applied” we mean that its *applicability* isn't the kind of thing I mean of a piece of wood when I say “I will be able to find many applications for it”. (PG: 319)

As a final remark, I would like to suggest that the approximation made by Stenlund between symbolic mathematics and formalism was the result of an overemphasis on the operational aspect of mathematical symbols to the detriment of their pictorial character as symbols¹⁸. While it is completely adequate to say that “a symbol is determined by how we operate with the sign for it” (Stenlund, 2015: 70) and that “it is the *operational* aspect of a symbol, its function in the calculus, its role in the manipulation and transformation of expressions, which *constitutes it as a symbol*” (*ibid*: 23), we must bear in mind that the manipulation of symbols in arithmetic is not an end in itself, neither a means to “win” in the arithmetical game, but a means to call our attention to certain *aspects* of the symbolism. The function of an arithmetical proposition, as Wittgenstein conceived it in his middle period, is to “draw our attention to a particular aspect of the matter” (*PR* X, §114*b*). And the function of an arithmetical proof is to display this aspect in the symbolism itself.

References

- Bell, E. T., 1937. *Men of Mathematics*. London: Victor Gollancz LTD.
- During, E., 2005. “Ni 'pure' ni 'appliquée': les usages de la géométrie chez Wittgenstein et Poincaré”, *Revue de métaphysique et de morale*, 46 (2), pp. 197-214.
- Frascolla, P., 1994. *Wittgenstein's Philosophy of Mathematics*. London, New York:

¹⁸ That this point is of crucial importance to an understanding of *Tractatus*'s views of arithmetic as well is well brought out by Secco and Noguez (2017).

Routledge.

- Frege, G., 1903. *Grundgesetze der Arithmetik*, Band II. Jena: Hermann Pohle.
- Klein, J., 1968. *Greek mathematical thought and the origin of algebra* (2nd ed. trans. by E. Brann, 1992). Cambridge: MIT Press.
- Medina, J., 2002. *The Unity of Wittgenstein's Philosophy: Necessity, Intelligibility, and Normativity*. Albany: State University of New York Press.
- Mühlhölzer, F., 2010. *Braucht die Mathematik eine Grundlegung? Ein Kommentar des Teils III von Wittgensteins Bemerkungen über die Grundlagen der Mathematik*. Frankfurt: Vittorio Klostermann.
- Narboux, J-P., 2001. "Aspects de l'arithmétique", *Archives de Philosophie*, 3, pp. 569-591.
- Rodych, V., 1997. "Wittgenstein on Mathematical Meaningfulness, Decidability, and Application", *Notre Dame Journal of Formal Logic*, 38 (2), pp. 195-225.
- Secco, G. D., Noguez, P. M. R., 2017. "Operar e exhibir: aspectos do conhecimento simbólico na filosofia Tractariana da matemática". *Revista Portuguesa de Filosofia*, 73 (3-4), pp. 1463-1492.
- Shapiro, S., 2000. *Thinking about Mathematics*. Oxford: Oxford University Press.
- Stenlund, S., 2015. "On the Origin of Symbolic Mathematics and Its Significance for Wittgenstein's Thought", *Nordic Wittgenstein Review* 4 (1), pp. 7-92.
- Wittgenstein, L., 1969. *Tractatus Logico-Philosophicus*, trans. D. F. Pears and B. F. McGuinness. London: Routledge & Kegan Paul.
- Wittgenstein, L., 1974 [PG]. *Philosophical Grammar*, R. Rhees (ed.), trans. A. J. P. Kenny. Oxford: Blackwell.
- Wittgenstein, L., 1975 [PR]. *Philosophical Remarks*, 2nd ed., R. Rhees (ed.), trans. R. Hargreaves and R. White. Oxford: Blackwell.
- Wittgenstein, L., 1978 [RFM]. *Remarks on the Foundations of Mathematics*, 3rd ed., G. H. von Wright, R. Rhees and G. E. M. Anscombe (eds.), trans. G. E. M. Anscombe. Oxford: Blackwell.
- Wittgenstein, L., 1979 [WWK]. *Ludwig Wittgenstein and the Viena Circle. Conversations Recorded by Friedrich Waismann*, B. McGuinness (ed.), trans. J. Schulte and B. McGuinness. Oxford: Blackwell.
- Wittgenstein, L., 1980 [LWL]. *Wittgenstein Lectures, Cambridge, 1930-1932: From the Notes of John King and Desmond Lee*, D. Lee (ed.). Totowa, New Jersey: Rowman and Littlefield.
- Wittgenstein, L., 1994 [Wi2]. *Wiener Ausgabe, Band 2. Philosophische Betrachtungen, Philosophische Bemerkungen*. Hg. von Michael Nedo. Wien, New York:

Springer.

Wittgenstein, L., 2012 [BT]. The Big Typescript: TS 213, German/English Scholar's Edition, trans. C. Grant Luchart and M. E. Aue. Oxford: Wiley-Blackwell.

CAUSALIDADE INTERDETERMINÍSTICA E FENÔMENOS QUÂNTICOS

MAYRA MOREIRA COSTA
UFMG

1. Introdução

O problema da aplicação da causalidade no domínio da mecânica quântica é caracterizado pela dificuldade em explicarmos causalmente determinados fenômenos preditos pelo formalismo da mecânica quântica de acordo com os modelos causais clássicos disponíveis. Neste trabalho, referir-me-ei especificamente ao fenômeno do contexto experimental do tipo EPR-Bohm, onde medições de propriedades de duas entidades quânticas espacialmente separadas exibem uma anticorrelação perfeita entre si e a dependência estatística entre elas não é anulada por uma causa comum, como ocorre no domínio clássico.

Como alternativa frente ao problema, defenderei a hipótese de que o domínio da causalidade não é restrito a fenômenos clássicos e que um modelo causal indeterminístico é uma alternativa frutífera para aplicação da causalidade a esse tipo de fenômeno quântico. Para fundamentar essa hipótese, seguirei os passos que se sucedem.

Primeiramente, apresentarei os modelos causais que podem ser obtidos a partir das teorias da causalidade de Hans Reichenbach (1944, 1956) e Wesley Salmon (1984, 1998) e explicarei por quais razões eles não poderiam ser aplicados ao domínio da mecânica quântica. Depois disso, oferecerei uma generalização do Princípio de Causa Comum (*PCC*) de Reichenbach e uma reformulação do seu Critério de Transmissão de Marca (*CTM*), somada a uma reformulação do modelo de bifurcação causal interativa (*BCI*) de Salmon e de sua teoria dos processos causais (1984, 1997, 1998). O modelo de bifurcação causal conjuntiva (*BCC*) de Reichenbach será mantido para a explicação dos fenômenos clássicos.

De acordo com essa alternativa, a diferença entre os dois tipos de

bifurcação causal fornecerá a chave para compreendermos a distinção entre probabilidade e causalidade na física clássica e na física quântica, um problema que desde o advento desta tem tirado o sono tanto de físicos, quanto de filósofos.

2. A teoria da causalidade de Reichenbach e o indeterminismo quântico

A teoria causal de Reichenbach (1944, 1956) é constituída por uma dupla redução: a redução da direção do tempo à direção da causalidade e a redução das relações causais a relações probabilísticas. O modelo que defendo nesta pesquisa, mantém a primeira redução, caracterizando-se como uma teoria causal do tempo, mas não mantém a segunda. Embora as correlações probabilísticas verificadas em distintos tipos de fenômeno ocupem um papel central na sua estruturação, as relações causais não serão entendidas como uma mera idealização de relações probabilísticas, como defendeu Reichenbach.

De acordo com a perspectiva de Reichenbach, o conceito de causalidade é atado à nossa capacidade de prever probabilisticamente eventos futuros:

[A] afirmação de que a natureza é governada por leis causais estritas significa que podemos prever o futuro com uma determinada probabilidade e que podemos impulsionar essa probabilidade próxima da certeza como desejarmos usando uma análise suficientemente elaborada do fenômeno sob consideração.
(REICHENBACH (1944): 2-3, tradução nossa).

Nossa capacidade de predição, no entanto, jamais seria inequívoca. Poderíamos nos aproximar da certeza, mas nunca a atingiríamos plenamente e, por isso, nunca verificaríamos de fato qualquer relação causal, apenas implicações probabilísticas.

Segundo Reichenbach, não podemos verificar qualquer lei física que não seja probabilística, mesmo entre as da física clássica (REICHENBACH (1944): 3). As leis causais estritas somente teriam significado por poderem ser interpretáveis sob a forma de uma lei probabilística equivalente e empiricamente verificável. Para ilustrar como se dá a interpretação de leis causais em termos de leis

probabilísticas, considere um exemplo adaptado de Reichenbach (1944, p. 2):

Zé Ninho mora na roça e está tentando derrubar uma goiaba determinada da árvore do vizinho com o seu estilingue. Aceitemos inicialmente que o disparo do estilingue causa o impacto da pedra em um determinado ponto, no caso, no galho que sustenta a goiaba. Considere A como as condições iniciais envolvidas no disparo da pedra, tal como o peso da pedra, a inclinação do elástico do estilingue e a distância em que Zé estende o elástico. Tais parâmetros podem ser medidos independentemente e de forma simultânea. Considere que B representa a descrição do ponto de impacto, no nosso exemplo o galho que sustenta a goiaba. Após experimentar vários casos de disparos de pedras com estilingue segundo aqueles parâmetros, Zé pode medi-los para um dado disparo e prever um ponto de impacto com certa probabilidade, obtendo uma implicação probabilística da forma “se A então B dada certa probabilidade”. Dessa relação, Zé também pode, por meio de uma extrapolação nunca segura, segundo Reichenbach, sustentar uma relação ideal, sob a forma de uma implicação estrita do tipo “se A' então B' ”, considerando os fenômenos ideais A' e B' . Assegurado pela sua grande habilidade, Zé Ninho poderia adicionar novos parâmetros à descrição do disparo, como, por exemplo, a força empregada ao se esticar o elástico, a influência da rotação da terra, a resistência do ar, a eminência de surgimento do vizinho etc, e isso tudo possibilitaria aumentar a probabilidade da implicação. No entanto, nunca verificaríamos Zé Ninho, nem qualquer outra pessoa ou grupo de pessoas, ainda que munidos de todo um aparato de medição avançado, conseguindo atingir a certeza, isto é, a garantia de uma implicação estrita, apesar de verificarmos um aumento de sua proximidade probabilística com a inclusão de parâmetros relevantes adicionais e medições mais precisas das condições iniciais dos fenômenos. Assim, de acordo com Reichenbach, a implicação ideal não verificável “se A' então B' ” só é provida de significado porque podemos substituí-la pela implicação

probabilística verificável “se A então probabilisticamente B”.

Para Reichenbach, um dos principais casos nos quais estão centrados os problemas filosóficos da mecânica quântica é exatamente em relação à transição de leis probabilísticas a leis causais estritas (1944, p.1). O diferencial no caso da física quântica em relação à física clássica seria o fato de não podermos, de forma alguma, atingir no primeiro domínio uma idealização semelhante a que podemos no segundo, não poderíamos formular leis causais que sejam interpretáveis nos termos de respectivas leis probabilísticas. A origem disso, de acordo com Reichenbach, está nas leis da mecânica quântica expressas pelo *princípio da incerteza* de Heisenberg (1925, 1927). Essas leis representam um limite de mensurabilidade: não podemos obter uma medida precisa de todos os parâmetros de estados físicos independentemente e de forma simultânea, mas somente de uma parte deles. Uma medição próxima da certeza para um parâmetro, acarreta em uma medição distante da certeza em outro.

Consideremos, para ilustrar, a relação de incerteza entre posição (q) e momento (p) de um elétron. *Grosso modo*, de acordo com esta relação, se obtivermos uma medida com acurácia da distribuição probabilística de q , a posição do elétron, não podemos obter com a mesma precisão a distribuição probabilística de p , seu momento. Não podemos medir tais propriedades independentemente e não podemos fazer predições que se aproximem da certeza em tal domínio, já que não há como medir valores simultâneos dos parâmetros envolvidos com exatidão arbitrária. Não poderíamos aqui, portanto, dar o passo que damos no exemplo da relação entre o disparo do estilingue e o ponto de impacto da pedra. Para Reichenbach, isso demonstra que temos apenas leis estatísticas no domínio da física quântica e que elas, de modo algum, podem ser convertidas em leis causais, mesmo que por meio de idealização, como ocorreria no caso da física clássica.

Podemos defender, não obstante, ao contrário de Reichenbach, que não há razões para pensarmos que causalidade seja redutível a nossa capacidade de predição e que o caráter indeterminístico da mecânica quântica antes de mostrar que não há causalidade em tal domínio, indica-nos a hipótese de que causalidade opera também sob mecanismos indeterminísticos.

Diretamente relacionado a esse ponto está o fato de que no domínio quântico as correlações probabilísticas do chamado *princípio de causa comum* (PCC), o princípio probabilístico fundamental da teoria da causalidade de

Reichenbach (1956), são violadas. Segundo o *PCC* de Reichenbach, se há uma correlação positiva entre dois eventos A e B , de modo que $P(A \& B) > P(A) \cdot P(B)$ e um não é causa do outro, então deve haver uma causa comum C , sendo que C compartimentaliza a relação estatística entre eles, de modo que os efeitos exibem uma independência estatística entre si. Quando A é estatisticamente independente de B , dado C (formalmente: $A \perp\!\!\!\perp B \mid C$), então as seguintes relações sustentam-se:

$$P(A \& B \mid C) = P(A \mid C) \cdot P(B \mid C) \quad (1)$$

$$P(A \mid B \& C) = P(A \mid C) \quad (2)$$

$$P(B \mid A \& C) = P(B \mid C) \quad (3)$$

Um conjunto de eventos que exiba tal arranjo e obedeça tais relações probabilísticas figura o que ele nomeou como *bifurcação causal conjuntiva* (*BCC*).

Fenômenos quânticos do tipo *EPR-Bohm*, por sua vez, não exibem correlações probabilísticas que constituiriam uma bifurcação causal conjuntiva. Eles exibem, no lugar, as correlações probabilísticas do que Salmon denominou como *bifurcação causal interativa* (*BCI*). Numa *BCI* a causa comum falha em compartimentalizar os seus efeitos, de modo que $P(A \& B \mid C) > P(A \mid C) \cdot P(B \mid C)$.

Na próxima seção, farei uma breve exposição da teoria da causalidade de Salmon, considerando a possibilidade das suas bifurcações interativas serem aplicáveis ao domínio indeterminístico.

3. A bifurcação causal interativa de Salmon

A teoria da causalidade de Salmon (1984, 1998) não reduz leis causais a leis probabilísticas, como a de Reichenbach. Sua ênfase, ao contrário, está voltada para os mecanismos causais entre processos, e não para correlações probabilísticas entre eventos.

Os dois conceitos centrais de sua teoria são: *interação causal* e *processo causal*. A interação causal entre processos é responsável pela produção de marcas em cada um deles. Uma marca pode ser entendida aqui como sinônima de sinal, informação ou registro. Um processo causal é algo que exhibe consistência de estrutura sobre o tempo e transmite marcas num intervalo espaço-temporal sob

o limite máximo da velocidade da luz. Uma marca é produzida instantaneamente no momento da interação e mantém-se inalterada durante o intervalo de sua propagação na ausência de uma interação local adicional.

Ainda que Salmon tenha formulado as correlações probabilísticas que se adequariam ao domínio quântico, através das bifurcações interativas, ele não considerava que a causalidade poderia ser aplicável a tal domínio, considerando as *BC*s próprias do domínio clássico juntamente com as *BCC*s de Reichenbach, além das bifurcações perfeitas, que seriam limites determinísticos de ambas. Suas razões para sustentar isso, todavia, não estavam diretamente conectadas ao caráter indeterminístico da mecânica quântica, mas antes ao fenômeno da *não-localidade* que entra em conflito com a sua teoria processual de transmissão de marca, já que sua teoria somente se adequaria a fenômenos quânticos, se fosse adequado interpretarmos tais casos sob a pressuposição da existência de variáveis ocultas locais. No entanto, desde a prova de Bell (1964, 1966) da impossibilidade de variáveis ocultas locais no domínio quântico, somada à demonstração empírica de que as predições da mecânica quântica estão corretas, esse tipo de suposição mostrou-se insustentável e em desacordo com os resultados experimentais.

Para compreendermos melhor esse ponto, vejamos uma representação gráfica simplificada de um arranjo experimental do tipo EPR-Bohm:

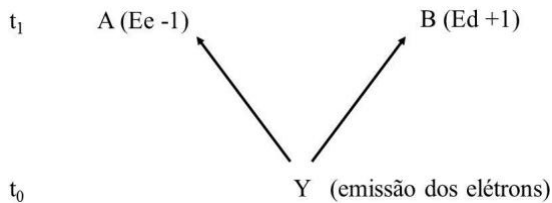


Fig.1. Dois elétrons em estado de singleto sendo emitidos de uma fonte comum.

No instante t_0 temos dois elétrons em estado de *singleto* sendo emitidos de uma fonte comum (Y). Suponhamos que um deles, *Ee*, se direcione para esquerda e o outro, *Ed*, para a direita. Em cada um dos pontos há um medidor, que mede o *spin* dos elétrons, seu *momento* angular intrínseco. Essa medida poderá ter um de dois valores rotacionais: *spin* “para cima” (+ 1) ou *spin* “para baixo” (- 1). No estado de

singleto, os elétrons estão correlacionados de tal forma que se o *spin* de um é para cima numa dada direção, o *spin* do outro será para baixo na mesma direção, eles exibem uma anticorrelação perfeita. De acordo com o formalismo da mecânica quântica, dada a emissão dos dois elétrons a partir de uma fonte comum, a probabilidade de ocorrer *A* e *B* é aleatória, de modo que podemos estipular que a probabilidade de *Ed* apresentar a propriedade +1 seria igual a 0,5 assim como a probabilidade do *Ee* apresentar a propriedade -1. Assim, $P(A|Y) = P(B|Y) = 0,5$ e a causa comum não torna seus efeitos estatisticamente independentes, de modo que $P(A|B \& Y) = 1 > P(A|Y) \cdot P(B|Y)$.

Segundo a teoria de Salmon, numa bifurcação interativa os efeitos ocorrem simultaneamente à causa comum, pois uma alteração de uma característica, no caso *A* e *B*, ocorreria em uma única intersecção local *Y* em t_0 e seria transmitida no intervalo $t_1 - t_0$ ao longo de cada ponto espaço-temporal. Isso só seria admissível no arranjo EPR-Bohm se interpretarmos o experimento sob a suposição da existência de variáveis ocultas locais, já que, dessa forma, as propriedades *A* e *B* já estariam definidas em t_0 . No entanto, se eliminarmos a suposição da existência de variáveis ocultas locais, *A* e *B* somente são determinadas num instante posterior t_1 . A exibição da propriedade -1 em *Ee* (*A*) ou da propriedade +1 em *Ed* (*B*) só estaria definida no momento da medição de algum dos dois em t_1 . Durante o intervalo $t_1 - t_0$ os estados das entidades quânticas são indefinidos. Desse modo, Salmon não pode dizer que a interação *Y* em t_0 produziu uma marca em cada um dos processos causais que se manteve inalterada até t_1 .

O modelo causal de bifurcação interativa de Salmon não pode, portanto, ser usado na explicação do experimento EPR-Bohm. Na próxima seção, apresentarei uma interpretação distinta das correlações da *BCI* sob uma nova formulação do conceito de processo causal aplicável a esse tipo de fenômeno. Antes disso, buscarei explicar por que as correlações das *BCI* não se adequam ao domínio clássico como pretendeu Salmon.

Considere que Zé Ninho é o campeão da barraca de tiro ao alvo duplo com estilingue da quermesse. Nesta barraca o desafio consiste em acertar dois alvos simultaneamente com a mesma goiaba. O jogo é arranjado de forma que os alvos estão posicionados no fundo da barraca, um do lado direito e um do lado esquerdo, e no centro mais a frente há um facão verticalmente colocado de modo que a intersecção da goiabada no ponto *x* com o facão, o evento *C*, corta ela ao

meio e faz com que a metade da direita acerte o alvo da direita, o evento *A*, e a metade da esquerda atinja o alvo da esquerda, o evento *B*. Desse modo, a probabilidade de acertar o alvo da direita, *A*, seria dependente da probabilidade de acertar o alvo da esquerda, *B*, e a causa comum *C* não os tornaria independentes, pois poderíamos estar certos de *A* dado *B*. Essa situação, no entanto, não é verificada na prática em qualquer experiência clássica, pelo contrário, o que verificamos corresponde às correlações da *BCC* de Reichenbach.

Em uma bifurcação conjuntiva, a causa comum torna os efeitos probabilisticamente independentes exatamente por que nada garante após uma interação que um dos co-efeitos seja predito com certeza dado o resultado observado de outro. No nosso exemplo, isso significaria que alguém que tivesse informação do aparato do jogo e de que Zé acertou somente o alvo da direita, não poderia afirmar com certeza se a outra metade da goiaba atingiu o alvo da esquerda, somente um exame da trajetória da metade da esquerda forneceria um resultado acurado. Não é difícil imaginarmos várias coisas que poderiam ter interferido no meio da trajetória impedindo a goiabada de atingir o alvo da esquerda por mais que tivéssemos a informação de que a outra metade atingiu o alvo da direita.¹ Claro que devemos considerar o efeito na interpretação de Salmon como imediatamente posterior à interação, e, por isso, como consistindo na mudança de direção do movimento dos objetos que se manteria até uma nova interação. Mas, desse modo, não observamos na prática qualquer processo causal transmitindo uma marca inalterada além do local de intersecção, já que interações adicionais entre os objetos ocorrem mesmo simultaneamente à intersecção. Observamos, por isso, as correlações da *BCC* de Reichenbach, pois nada garante, após uma interação causal, que qualquer uma das marcas provocadas em cada um dos processos se mantenham correlacionadas e inalteradas ao longo de suas trajetórias. Somente para objetos clássicos idealmente isolados ou para fenômenos quânticos que podemos atribuir as correlações das *BCI*, de forma que só as verificamos no segundo caso.

Na próxima seção, apresentarei uma sugestão de teoria causal do tempo, a partir de reformulações das ideias de Salmon e Reichenbach, com o propósito

¹ Russell, no seu artigo “On the notion of cause” (1913), já chama a atenção para esse ponto (ainda que não especificamente sobre esse caso e nesse tipo de cenário) ao tratar do equívoco da contiguidade espaço-temporal entre causa e efeito.

de que ela abranja também o tipo de fenômeno quântico correspondente às correlações probabilísticas das *BCI*. Essa reformulação repousará, sobretudo, numa concepção distinta da concepção salmoniana de processo causal e em generalizações do Princípio de Causa Comum (*PCC*) e do Critério de Transmissão de Marca (*CTM*) de Reichenbach.

4. Um modelo causal indeterminístico

Tendo em mente a estratégia de elaboração de um modelo causal que figure as correlações das *BCI*s para sua aplicação a fenômenos quânticos e as correlações das *BCC*s para fenômenos clássicos, podemos sugerir uma generalização disjuntiva do *PCC* de Reichenbach:

*PCC**: Se há uma correlação positiva entre dois eventos *A* e *B*, de modo que $P(A \& B) > P(A) \cdot P(B)$ e *um não é causa do outro*, então deve haver uma causa comum, um evento *C*, sendo que *ou* *C* compartimentaliza a relação estatística entre eles, de modo que $P(A \& B|C) = P(A|C) \cdot P(B|C)$, constituindo uma *BCC* *ou* *C* não compartimentaliza a relação estatística entre eles, de modo que $P(A \& B|C) > P(A|C) \cdot P(B|C)$, constituindo uma *BCI*.

A exigência prévia condicional de que *um não seja causa do outro* será dada pela reformulação de um princípio mais fundamental também originalmente proposto por Reichenbach (1928): o Critério de Transmissão de Marca (*CTM*). Considerando *E*₁ e *E*₂ como dois eventos distintos, de acordo com o *CTM*:

*Se E*₁ *é causa de E*₂, *então uma pequena variação* (uma marca) *em E*₁ *é associada a uma pequena variação em E*₂, *ao passo que pequenas variações em E*₂ *não são associadas a variações em E*₁ (REICHENBACH, 1958 (1928), p. 136, tradução nossa, grifo do autor).

O *CTM* não pressupõe uma direção temporal, fornecendo um critério independente para a determinação da direção da causalidade. Por isso, um modelo causal

sustentado em tal critério, somado à hipótese de que a direção do tempo é determinada pela direção da causalidade, caracteriza-se como um tipo de teoria causal do tempo. Segundo o *CTM* de Reichenbach, o que distinguiria um evento como causa de outro seria a possibilidade de uma pequena variação no primeiro ser associada a uma pequena variação no segundo e não *vice-versa*. O primeiro evento chamaremos de causa e o segundo de efeito, e, aplicando a redução causal da direção do tempo, a causa chamaremos de anterior e o efeito de posterior

O termo “ser associado” tem o problema de deixar muito vago qual tipo de associação unidirecionada entre eventos seria essa a qual recorre Reichenbach. Um indício de resposta advém do fato do envio de informação ser na experiência temporalmente direcionado, no sentido em que sinais não podem ser enviados na direção de eventos posteriores para anteriores. É fácil perceber uma consequência disso no fato de não podemos manipular o passado. Somado a isso, devemos considerar também os postulados da relatividade e, devido ao segundo, o fato de que sinais são enviados sob o limite de velocidade da luz. Podemos então reformular o *CTM* do seguinte modo:

*CTM**: Se E_1 é causa de E_2 , então uma informação de E_1 é transmitida sob o limite da velocidade da luz para E_2 , ao passo que informações de E_2 não podem ser transmitidas para E_1 .

Para ilustrar essa formulação em um exemplo cotidiano, considere o seguinte. Zé Ninho marcou um encontro com Maria Ninho no dia da festa de São João, mas esqueceu de combinar o local. Zé Ninho chega uma hora mais cedo na festa e resolve escrever um correio elegante para Maria, escrevendo no papel o local do encontro, chamemos esse evento de E_1 . Quando Maria Ninho chega na festa, uma hora depois, a menina do correio elegante entrega para ela o papel com a informação que Zé escreveu nele, chamemos esse evento de E_2 , de forma que uma informação é enviada de E_1 para E_2 . Zé ao chegar na festa, em E_1 , não pode receber nesse instante uma informação que teria sido enviada por Maria no instante de E_2 . Desse modo, concluímos que E_1 é causa de E_2 , pois uma informação em E_1 é transmitida sob o limite da velocidade da luz para E_2 , ao passo que informações em E_2 não podem ser transmitidas para E_1 . Podemos aplicar o mesmo princípio sem nenhuma dificuldade em qualquer outra experiência

cotidiana ou científica de envio de informação, que sempre é direcionada.²

É por conta da assimetria da informação que não esperamos que informações viajem no tempo em direção ao passado, mas só em direção ao futuro. Isto é, qualquer informação que chegue até nós, viaja em uma direção temporal determinada, do passado em direção ao futuro, de eventos anteriores em direção a eventos posteriores. A radiação com a qual entramos em contato hoje proveniente de galáxias distantes, por exemplo, viajou de um momento do passado de tais galáxias até o nosso presente. Não entramos em contato com radiações futuras delas, e de modo geral, não interagimos com sinais provenientes do futuro. Outra vantagem dessa formulação é que ela também é válida no domínio quântico, já que não podemos, mesmo nesse domínio, transmitir informações em uma direção determinada do tempo, correspondente ao que chamamos de passado. Somente podemos transmitir informações em direção ao futuro, do mesmo modo que no domínio clássico.

Com essa formulação do *CTM*, nós podemos voltar ao problema da interpretação causal das *BC/s* para o domínio quântico e explicitar qual a diferença na concepção de processo causal entre o modelo sugerido e a teoria salmoniana. Os mecanismos causais da teoria de Salmon estão centrados, sobretudo, no *CTM* introduzido por Reichenbach, mas sob uma interpretação essencialmente processualista:

Uma marca, que foi introduzida em um processo por meio de uma única intervenção no ponto A, é transmitida para o ponto B se, e somente se, ela ocorre no ponto B e em todos os estágios do processo entre A e B na ausência de intervenções adicionais (SALMON, 1998, p. 197).

² Para uma demonstração em linguagem mais científica da impossibilidade de envio bidirecional de informação, veja “An Inquiry into the Possibility of Nonlocal Quantum Communication” (2015) de John Cramer e Nick Herbert. Ainda sobre esse ponto, mas de um modo mais geral e com foco em problemas centrais de filosofia da física, o artigo “Understanding the Time-Asymmetry of Radiation” (2013) de Jill North é bastante esclarecedor.

Se não consideramos que os processos causais são responsáveis pela transmissão de uma marca que se mantém inalterada durante um intervalo espaço-temporal, como não o fazemos no *CTM**, então podemos aplicá-lo aos fenômenos do tipo *EPR-Bohm*, bem como aos clássicos, pois, como vimos, a exigência de Salmon de que uma marca se mantenha inalterada após uma interação sequer pode ser verificada no domínio clássico. A distinção entre os dois tipos de fenômeno envolvendo uma causa comum poderá ser dada pelo *PCC**, que nos permite classificar o tipo de bifurcação causal, dada as correlações probabilísticas verificadas.

Voltando ao nosso exemplo simplificado do arranjo experimental *EPR-Bohm*, aplicando o *CTM**, poderemos concluir que o evento *Y* é causa comum de *A* e *B*, pois nenhuma informação pode ser enviada na direção dos eventos de medição para o evento de interação em *Y*, nem qualquer informação pode ser enviada simultaneamente na direção do evento de medição de *E_e*, o evento *A*, para o evento de medição em *E_d*, o evento *B*, ou vice-versa. A direção da causalidade só é definida no instante de medição, pois um processo causal só é completo quando há transmissão de informação entre dois eventos.

Tal concepção, menos comprometida e mais generalizada da transmissão de marca, acarreta também, por isso, em uma concepção distinta de *processo causal*. Salmon define um processo como qualquer coisa que exiba consistência de estrutura sobre o tempo (1984, p. 144). Em sua versão mais recente envolvendo quantidades conservadas, Salmon define um processo causal como uma linha do universo de um objeto que transmite uma quantidade conservada em cada ponto espaço temporal de sua trajetória (1997, pp. 462, 468). Se não atarmos os processos causais a tal condição salmoniana de transmissão localizada, então um processo causal poderá ser definido como *um conjunto de eventos causalmente conectados por uma informação transmitida sob o limite da velocidade da luz*.

A noção de evento e o *CTM**, portanto, serão mais fundamentais nessa reformulação do que o conceito de processo causal, que é estabelecido a partir deles. Temos, por isso, de definir um evento sem pressupor previamente uma direção temporal, já que isso deve ser dado de forma independente pelo *CTM**. Podemos recorrer aqui à definição de mudança de Russell (RUSSELL, 1903, p. 469), que para nós servirá para estabelecer a ocorrência de um evento: um evento ocorre quando há diferença entre os valores de verdade de proposições referentes

a uma propriedade y de uma entidade x , sendo que a única diferença entre as proposições se refere ao instante no tempo. Por exemplo, se a proposição expressa pela frase “A goiaba caiu da árvore” é falsa em um momento t , mas é verdadeira em um outro momento t' , então podemos concluir que uma mudança ocorreu com a Goiaba, e, portanto, um evento.

Definir um evento como a ocorrência de uma mudança de uma propriedade de uma entidade, que, por sua vez é definida pela diferença de valores de verdade entre proposições referentes a diferentes instantes do tempo, pode levar alguém a objetar que o *CTM** fundamentado em tal definição é circular por recorrer ao conceito de instantes do tempo com o objetivo de fornecer sua direção. No entanto, somente com a definição de mudança de Russell, não pressupomos nenhuma direção temporal, afinal, ela não define t como anterior a t' , nem vice-versa, apenas que houve uma mudança com uma entidade, mas sem qualquer direcionamento. Somente ao aplicarmos o *CTM** é que podemos definir eventos como anteriores ou posteriores entre si, dada a direção do envio de informação.

Além disso, temos uma motivação adicional para desconsiderar a formulação salmoniana de transmissão de marca: ela pressupõe o absolutismo temporal, ainda que fosse intencionada a alicerçar de forma relacionista a natureza do tempo. De acordo com a posição relacionista sobre a natureza do tempo, não há fluxo do tempo, ou qualquer coisa correspondente a uma forma substancial do tempo, que passe independente da ocorrência de mudanças nos objetos. Para um absolutista do tipo substancialista, ao contrário, há passagem do tempo ainda que nenhuma mudança ocorra com objetos. Isso significa que se o universo todo parasse de mudar a partir de agora, por exemplo, para o relacionista o tempo de todo o universo terá parado, mas para o absolutista somente as coisas parariam de mudar, o tempo continuaria passando. De acordo com a formulação de Salmon, haveria passagem de tempo na ausência de qualquer intervenção, uma marca seria definida por uma interação e se manteria intacta, desde que não sofresse outra interação que provocasse alguma alteração adicional no processo. Isso significa afirmar que o tempo passaria ainda que o processo não sofra qualquer mudança, o que pressupõe uma concepção substancialista do tempo.

Se considerarmos os resultados da teoria da relatividade, aplicando a visão relacionista do tempo a cada sistema de referência, teremos de concluir que após a ocorrência de uma mudança (uma marca) em um determinado objeto

(sistema referencial) não há passagem de tempo até que ocorra um novo evento de mudança. Desse modo, até a ocorrência de outro evento com elétron não há um intervalo espaço- temporal definido. Isso é totalmente compatível com a formulação CTM*, já que para determinados a direção da causalidade e, portanto, a direção do tempo, é necessário que pelo menos dois eventos de mudança estejam conectados por um sinal: um processo causal só é completo quando um sinal é completamente enviado entre um evento e outro. Não há processos causais definidos na ausência de interação adicional.

Ademais, supor que a direção do tempo só é definida no instante da medição, evita que sejamos levados a conclusões paradoxais, como os famosos paradoxos de Zenão, que resultam apenas de uma concepção absoluta do tempo, na acepção segundo a qual poderíamos subdividi-lo do mesmo modo que uma série ordenada numérica puramente abstrata. Evitamos tais paradoxos se compreendermos o tempo dos processos como inteiramente resultantes e dependentes da ocorrência direcionada de interações.

5. Conclusão

O objetivo central deste ensaio foi a apresentação de um esboço de um modelo causal aplicável ao domínio indeterminístico da física quântica. Além da análise de modelos causais de grande relevância para a discussão, ofereci uma breve introdução ao problema da aplicação da causalidade no domínio quântico, relacionando-o aos problemas incidentes da direção do tempo e da causalidade. Tal alternativa fornece também uma explicação para a distinção entre os mecanismos causais e as correlações probabilísticas inerentes aos domínios quântico e clássico.

Em seu artigo "On the notion of cause" (1913), Bertrand Russell defendeu que a noção de causalidade não passava de uma relíquia de tempos antigos, ainda sustentada pelos filósofos da sua época pelo fato deles não perceberem qual seria o problema nisso. O problema, de acordo com ele, estaria em uma ligação inextrincável entre a noção de causalidade e associações incorretas, de forma que seria desejável excluí-la completamente do vocabulário filosófico (RUSSELL, 2003 (1913), p. 164). Entre as associações equivocadas apontadas por Russell está exatamente a conexão da noção de causalidade com a noção de determinismo, e,

nesse ponto, penso que devemos concordar que a noção de causalidade considerada por ele é bastante questionável. Ainda assim, podemos argumentar que essa ligação entre causalidade e determinismo não é inextrincável como ele defendeu e que não só é totalmente plausível a defesa de um modelo não determinístico, como completamente inadequado reduzir um termo ao outro.

Desse modo, creio que o termo “lei da causalidade”, presente na famosa citação de Russell a favor do ostracismo da noção de causa, seria proveitosamente substituída pelo termo “determinismo causal”: o *determinismo causal*, como muito do que se passa entre os filósofos, é uma relíquia de um tempo antigo, sobrevivendo, como a monarquia britânica, somente porque é erroneamente suposto não causar danos.

Referências bibliográficas

- Bell, J. (1964). “On the Einstein-Podolsky-Rosen Paradox”, *Physics* 1, 195-200, 1964.
- Bell, J. (1966). “On the Problem of Hidden Variables in Quantum Mechanics”, *Reviews of Modern Physics* 38, pp. 447-75, 1966.
- Cramer, J. e Herbert, N. (2015). “An Inquiry into the Possibility of Nonlocal Quantum Communication”, *arXiv, Quantum Physics*, arXiv: 1409.5098v2.
- Einstein, A. Podolsky, B. e Rosen, N. (1935). “Can quantum-mechanical description of reality be considered complete?”, *Physical Review*, v. 47, p. 777-80, 1935.
- Heisenberg, W. (1925). “Über quantentheoretische Umdeutung kinematischer und mechanischer Beziehungen”, *Z. Phys.* 33, 879-893, 1925.
- Heisenberg, W. (1927) “The physical content of quantum kinematics and mechanics”, trad. in: Wheeler, J. A. & Zurek, W. H. (ed.). *Quantum theory and measurement*, Princeton: Princeton University Press, p. 62-84, 1983.
- North, J. (2003). “Understanding the Time-Asymmetry of Radiation”, *Philosophy of Science (Proceedings)*, vol. 70, December 2003, pp. 1086-1097
- Reichenbach, H. (1928). *The Philosophy of Space and Time*, trad. M. Reichenbach, New York: Dover, 1958.
- Reichenbach, H. (1944). *Philosophic Foundations of Quantum Mechanics*, Mineola, New York: Dover Publications, 1998.
- Reichenbach, H. (1956). *The Direction of Time*, Berkeley, Los Angeles, London: University of California Press, 1971.

- Russell, B. (1903). *Principles of Mathematics*, New York – London: W. W. Norton and Company.
- Russell, B. (1913). “On the notion of cause”, in *Russell on Metaphysics: selections from the writings of Bertrand Russell*, edit. Stephen Munford, London – New York: Routledge, 2003.
- Salmon, W. (1984). *Scientific Explanation and the Causal Structure of the World*, Princeton: Princeton University Press, 1984.
- Salmon, W. (1997). “Causality and Explanation: A Reply to Two Critiques”, *Philosophy of Science*, 64: 461-77
- Salmon, W. (1998). *Causality and Explanation*, Oxford: Oxford University Press.



Editora
UFPel

DISSERTATIO
FILOSOFIA