Vol. IV, No. 34, December 2012

Disputatio

International Journal of Philosophy

Special Issue

Logic, Norms and Ontology. Recent Essays in Luso-Brazilian Analytic Philosophy

Edited by João Branquinho and Guido Imaguire

Centro de Filosofia da Universidade de Lisboa www.disputatio.com

Disputatio aims at publishing first-rate articles and discussion notes on all aspects of analytical philosophy, but especially those dealing with current issues in the philosophies of language, logic, and mind, and also in epistemology and metaphysics, written in English or Portuguese. The discussion notes need not be on a paper originally published in our journal. Articles of a purely exegetical or historical character will not be published. Although book reviews are usually commissioned, we also welcome proposals for reviews of recent books.

Disputatio practices double blind refereeing, and follows the 'Guidelines for the Handling of Manuscripts by the Editors of Philosophy Journals' of the Association of Philosophy Journal Editors.

Disputatio also accepts proposals for special issues of edited volumes on specific relevant topics or philosophical problems. A detailed proposal is to be made to the editors of the journal, including a brief description of the topic, and why it is relevant, information about the editor or editors of the special issue, and of the contributors. The proposal should also contain an abstract or summary of the papers to be included. The editors of the special number of the journal must guarantee that the papers included are suitably refereed. The editorial committee has the final word on the acceptance of the included articles.

Manuscripts with no more than 7,000 words should be typed, and will not be returned. They should also be prepared for anonymous refereeing. Authors are encouraged to submit articles by e-mail (please quote your paper's title). If submitting by regular mail authors must enclose an electronic version. Authors of accepted articles will be asked to prepare a final electronic version in accordance with *Disputatio's* style.

Manuscripts are accepted for review under the understanding that the work submitted for publication has not been published elsewhere, that it is not under consideration for publication elsewhere, and that it will not be submitted for publication elsewhere unless the editors of *Disputatio* have rejected it.

Disputatio is published twice a year, in November and May.

Submissions and mail are to be sent to the managing editor: Célia Teixeira, Centro de Filosofia da Universidade de Lisboa, Faculdade de Letras, Alameda da Universidade, 1600-214 Lisboa, Portugal. E-mail: celia.teixeira@gmail.com. Publishers should send review copies to this address.

PUBLISHED BY	SPONSORS
VNIVERSITAS OLISIPONENSIS **Centro de Filosofia** Universidade de Lisboa	**FCT** **Fundação para a Ciência e a Tecnologia** MINISTÉRIO DA CIÊNCIA E DA TECNOLOGIA Apoio do Programa Operacional Ciência, Tecnologia, Inovação do Quadro Comunitário de Apoio III.

Directores: João Branquinho e Teresa Marques. Publicação semestral. N.º de registo no ICS: 120449. NIPC: 154155470. Sede da redacção: Centro de Filosofia, Faculdade de Letras de Lisboa, Alameda da Universidade, 1600-214 Lisboa.

Disputatio

International Journal of Philosophy

Vol. IV, No. 34, December 2012

Published by Centro de Filosofia da Universidade de Lisboa
ISSN: 0873 626X — Depósito legal n.º 106 333/96

Table of Contents

Contributors

J.R. Arenhart, Federal University of Fronteira do Sul

João Branquinho, University of Lisbon and LanCog Group (UL)

Cláudio F. Costa, Federal University of Rio Grande do Norte

Fernando Ferreira, University of Lisbon

Pedro Galvão, University of Lisbon and LanCog Group (UL)

Nelson G. Gomes, University of Brasília

Dirk Greimann, Fluminense Federal University

E. Hermann Haeusler, Pontifical Catholic University of Rio de Janeiro

Breno Hax Jr., Federal University of Paraná and LanCog Group (UL)

Guido Imaguire, Federal University of Rio de Janeiro and LanCog Group (UL)

D. Krause, Federal University of Santa Catarina

André Leclerc, Federal University of Ceará

António Lopes, University of Lisbon and LanCog Group (UL)

Luiz Carlos Pereira, Pontifical Catholic University of Rio de Janeiro

André Porto, Federal University of Goiânia

Marco Ruffino, Federal University of Rio de Janeiro and LanCog Group (UL)

Ricardo Santos, University of Évora and LanCog Group (UL)

Rui Silva, University of the Azores and LanCog Group (UL)

Célia Teixeira, University of Lisbon and LanCog Group (UL)

Paulo A. S. Veloso, Federal University of Rio de Janeiro

Introductory Note

The present special issue of *Disputatio* brings together some of the best work recently done in Brazil and Portugal in the tradition of analytic philosophy (broadly conceived). Over the past ten years or so we have witnessed an impressive growth of analytic philosophy in both countries, either in terms of quantity or in terms of quality of the produced philosophy. We hope that this volume capture, at least partly, the dynamics and strength of such development. The range of philosophical problems and topics covered by the contributed essays is vast, cutting across several philosophical disciplines. Indeed, one can find therein issues in philosophical logic, meta-philosophy, ethics, aesthetics, philosophy of science, philosophy of language, philosophy of mathematics and metaphysics. Such variety of subject-matter is also a trait of recent Luso-Brazilian analytic philosophy.

The Editors
João Branquinho
Guido Imaguire

Uma Lógica da Indistinguibilidade

J.R. Arenhart
Universidade Federal da Fronteira Sul

D. Krause
Universidade Federal de Santa Catarina

BIBLID [0873-626X (2012) 34; pp. 555-573]

1 Introdução

Segundo uma concepção bastante em voga atualmente, a ontologia com a qual devemos nos comprometer deve de algum modo ser 'inferida' de nossas melhores teorias científicas, ou seja, dito de modo resumido, se desejamos saber o que há, devemos indagar às nossas melhores teorias. Esta posição, que é uma forma de naturalismo em ontologia, encontra algumas dificuldades no caso da mecânica quântica não-relativista. Segundo certas interpretações, o formalismo desta teoria é compatível com pelo menos dois tipos bastante distintos de ontologias: uma ontologia 'clássica' na qual as entidades tratadas por esta teoria são vistas como *indivíduos* em alguma acepção,[1] ou alternativamente, também é possível defender que a teoria nos

[1] Porém, podendo ser indiscerníveis, alguma versão do Princípio da Identidade dos Indiscerníveis (PII), de Leibniz, seria supostamente violado por estas partículas, pelo menos segundo alguns autores (algumas referências sobre este tópico são apontadas em Muller & Seevinck [2009], embora esses autores não concordem com a violação do PII pela mecânica quântica). O princípio nos garante que se objetos são numericamente distintos então deve existir alguma propriedade ou relação que os diferencie. Neste caso, se assumimos que as partículas quânticas são indivíduos, então o princípio de individualidade para elas não pode se basear em PII. Recentemente, no entanto, alguns autores tentaram estabelecer a validade de uma forma do PII neste contexto (ver Saunders [2006], Saunders e Muller [2008], Muller e Seevinck [2009]). No entanto, em nenhum desses trabalhos o conceito de "propriedade" é dado de forma rigorosa.

compromete com uma ontologia de *não-indivíduos*,[2] objetos para os quais as leis da identidade, conforme formuladas na lógica clássica, não se aplicam irrestritamente (ver French e Krause [2006] cap. 4 e 6). O problema é que a contraparte matemática da teoria não nos dá indicações que permitam favorecer uma destas interpretações em detrimento da outra. Assim, aparentemente, no caso desta teoria, a disputa terá que ser travada no campo da argumentação filosófica.

Adotemos uma terminologia provisória e informal. Um indivíduo será aqui (redundantemente) entendido como uma entidade (por falta de um termo melhor) que possa pelo menos em princípio ser identificado de modo inequívoco em qualquer contexto, em especial, sendo diferente de todas as demais entidades que forem dele distintas. Por exemplo, se supusermos a teoria de conjuntos Zermelo-Fraenkel com o axioma da escolha (ZFC), então todo conjunto pode ser bem-ordenado,[3] em particular o conjunto $\mathfrak{R}$ dos números reais é bem ordenado. O que acontece é que, como se sabe, não podemos exprimir essa boa ordem por uma fórmula da linguagem de ZFC. Em particular, como $(0,1) \subseteq \mathfrak{R}$, este subconjunto terá um menor elemento, que, no entanto, não pode ser expresso na linguagem (por exemplo, dando-lhe um nome). Porém, mesmo assim ele é um *indivíduo* na nossa acepção, sendo distinto de qualquer outro número real, como implica a lógica clássica, que subjaz à teoria usual dos números reais.

Por um *não-indivíduo*, entenderemos qualquer entidade que não obedece as condições acima. Não-indivíduos podem ser de diferentes 'espécies', e quando são *indiscerníveis*, ainda que possam ser co-

[2] O termo *não-indivíduo*, apesar de não muito adequado, é o padrão nestas discussões. É utilizado para designar objetos que violam alguma forma do princípio de individuação, em particular, para objetos que violam as chamadas leis da identidade, e mais especificamente a propriedade reflexiva da identidade, ou seja, para qualquer objeto x, x=x. Isso se deveria não ao fato de haver objetos que não são idênticos a eles mesmos, mas à circunstância de que a noção de identidade não poderia ser aplicada. Ver French & Krause [2006] para uma discussão pormenorizada. Outro conceito importante e não muito claro nessas discussões é o de "lógica clássica". Voltaremos a essa questão mais abaixo.

[3] Uma boa ordem sobre um dado conjunto é uma ordem parcial (reflexiva, anti-simétrica e transitiva) relativamente à qual todo subconjunto do conjunto dado tem menor elemento (um elemento do subconjunto que é menor—na ordem dada—que todos os outros elementos do subconjunto).

ligidos em coleções da vários deles, por definição não poderão ser discernidos por quaisquer critérios que se imagine, nem mesmo em princípio. Um não-indivíduo pode aparecer em um contexto e, se de algum modo for substituído por algum outro de mesma espécie, o contexto não se altera. Obviamente, uma coleção de não-indivíduos não satisfaria o axioma da extensionalidade de uma teoria de conjuntos como ZFC. Claro que essas caracterizações são imprecisas, e suas definições rigorosas dependerão da linguagem e da lógica empregadas. Deste modo, não há uma única definição de indivíduo, e o mesmo se dá para a de não-indivíduo, mas isso não impede que busquemos tornar estes conceitos mais precisos utilizando algum aparato lógico.

Com efeito, um modo de concebermos um pouco mais rigorosamente o que são indivíduos, ainda que haja dificuldade de se caracterizar precisamente o que é a lógica clássica,[4] seria sustentarmos que um indivíduo obedece as regras da teoria clássica da identidade, de primeira ordem ou de ordem superior (ou de alguma teoria de conjuntos calcada na lógica clássica). Assim, em especial, um não-indivíduo não obedece a *auto-identidade*, isto é, a lei fundamental $a=a$. Isso no entanto, segundo nossa caracterização, não implica que $a\neq a$, mas que a noção de identidade *não se aplica* aos não-indivíduos (o que está de acordo com a posição de Erwin Schrödinger, por exemplo, em Schrödinger [1952], pp.17-18; ver French & Krause [2006]). Esta hipótese parece essencial para podermos caracterizar não-indivíduos. Com efeito, admitamos que eles podem estar relacionados por uma relação mais fraca de `indiscernibilidade', que representaremos por '$\equiv$'. Assim, se supusermos, como parece razoável, que se $x\equiv y$ então eles são inter-substituíveis *salva veritate*, ou seja, vale algo como o axioma da substitutividade da igualdade da lógica de primeira ordem (a saber, algo como $x\equiv y \rightarrow (\alpha(x) \rightarrow \alpha(y))$ --ver mais à frente), se assumirmos além disso que $\forall x(x\equiv x)$, restituiríamos os

[4] De modo geral, podemos dizer que por lógica clássica entendemos o cálculo usual de predicados de primeira ordem com ou sem igualdade (conforme axiomatizado, por exemplo, em Mendelson [1987]), ou alguns de seus subsistemas, como o cálculo proposicional clássico, ou mesmo sistemas de 'grande lógica' como as lógicas usuais de ordem superior ou as teorias usuais de conjuntos, como ZFC e mesmo a teoria de categorias como usualmente concebida.

axiomas da igualdade da lógica elementar clássica, e então em nada estaríamos contribuindo efetivamente, mas apenas mudando a notação da igualdade para '≡'. Portanto, tendo-se em vista a discussão sobre a ontologia associada à MQ não-relativista, resulta que uma das dificuldades que surgem é que a lógica clássica, que é a lógica subjacente à mecânica quântica não-relativista, não parece compatível com uma ontologia de não-indivíduos, entendidos no sentido acima, pois a identidade sempre faz sentido para todas as entidades tratadas pela lógica clássica. Assim, se desejamos sustentar que uma ontologia de não-indivíduos é plausível, pelo menos no caso desta teoria em particular, parece razoável buscarmos por sistemas de lógica que nos permitam tratar de objetos para os quais a identidade e diferença não façam sentido.

Este tipo de investigação é importante filosoficamente, como atestam muitos autores, que sustentam que esta ontologia, a de não-indivíduos, é a mais natural para uma possível interpretação da mecânica quântica, principalmente se levarmos em conta determinadas interpretações de seus aspectos experimentais (ver a discussão histórica em French e Krause [2006] cap. 3). Isto torna imperiosa uma busca por mais rigor nos termos nos quais se expressam os conceitos de tal ontologia. Não entraremos aqui nas discussões sobre qual ontologia é mais adequada neste caso, nos restringindo aos problemas dos fundamentos lógicos da ontologia de não-indivíduos, que caracterizaremos abaixo.

Neste trabalho, apresentaremos uma linguagem formal de primeira ordem cujo objetivo principal é permitir o tratamento rigoroso de objetos para os quais a identidade e a diferença não se apliquem, ou seja, linguagens que possam tratar de não-indivíduos nesta acepção do termo. Este tipo de lógica é conhecido na literatura como *não-reflexiva* ou *para-reflexiva* (por exemplo, da Costa e Krause [1994], [1997] da Costa e Bueno [2009]). Vamos também considerar alguns problemas que surgem quando desejamos, ao mesmo tempo em que adotamos uma ontologia de não-indivíduos fazendo uso das linguagens por nós propostas, utilizar alguma das teorias de conjuntos clássicas[5] como metalinguagem para se estabelecer a semântica destas

[5] Grosso modo, estamos considerando como clássica uma teoria de conjuntos cuja lógica subjacente é a lógica clássica.

linguagens. Aqui, estamos supondo que a ontologia com a qual nos compromete uma teoria é determinada pelas entidades que devem existir no domínio de quantificação para que as sentenças da teoria sejam verdadeiras, no sentido tarskiano. Como veremos, se o domínio de quantificação dado por uma estrutura na qual interpretamos uma teoria formulada utilizando-se o sistema de lógica aqui proposto for um conjunto no sentido clássico, então as entidades com a qual trata a teoria serão indivíduos, violando, de certo modo, a motivação para se utilizar a lógica proposta neste trabalho.

A linguagem que apresentaremos é uma pequena modificação de uma proposta feita primeiramente por Newton C. A. da Costa em seu livro *Ensaio sobre os Fundamentos da Lógica*, (da Costa [2008] pp. 138-141), e constitui-se basicamente de uma linguagem bissortida de primeira ordem com certas restrições que comentaremos adiante, dando origem a um sistema que foi por ele batizado de lógica de Schrödinger. O objetivo que tinha em mente ao apresentar esta lógica era mostrar que o princípio de identidade, conforme formalizado por certa formulação[6], pode ser derrogado, ou seja, pode-se conceber um sistema de lógica, a lógica de Schrödinger, no qual a identidade ou diferença não se aplique a todas as entidades com as quais se pretende tratar. Este sistema de lógica era também motivado pelas dificuldades em se tratar da identidade e diferença quando falamos de partículas elementares, e deveria refletir certas intuições de E. Schrödinger quem, como dito acima, ao falar sobre partículas elementares da física quântica, insistia em que a questão sobre sua identidade ou diferença, em certos contextos, não faria sentido. Ou seja, (apesar de nem Schrödinger nem da Costa utilizarem esta terminologia), estes objetos são certo tipo de não-indivíduos: "Está além da dúvida que a questão da 'igualdade', da identidade [no que concerne as partículas elementares], real e verdadeiramente não tem sentido." (Schrödinger [1952] p. 18).

[6] da Costa utilizava a propriedade reflexiva da identidade da lógica clássica de primeira ordem, ou seja, $\forall x(x=x)$, como uma forma de representar o princípio da identidade.

2 Uma lógica para a indistinguibilidade

Adotaremos em nossa exposição a idéia original de da Costa para restringir a identidade de modo que faça sentido apenas para alguns objetos do domínio do discurso. Para fazer com que a identidade no sentido usual não se aplique a certas entidades, da Costa empregou uma linguagem bissortida, com duas espécies de termos individuais, e uma mudança na definição de fórmulas. Uma das espécies de termos, que podemos supor que seja a primeira, denotaria objetos microscópicos, e a outra, a segunda, denotaria as entidades macroscópicas. A restrição feita na definição de fórmula é a de impedir que o símbolo de identidade seja uma fórmula quando ladeado por pelo menos um termo de primeira espécie. Os axiomas da lógica clássica, observadas as diferenças de termos, completavam a apresentação. Em nosso caso, para formularmos uma lógica da indistinguibilidade, acrescentaremos ainda a esta linguagem uma relação binária de indistinguibilidade com postulados adequados para os objetos de primeira espécie. Os objetos de segunda espécie também poderão relacionar-se pela relação de indistinguibilidade, mas neste caso esta relação colapsará na identidade para estes objetos.

Para vermos como esta idéia funciona, apresentaremos agora uma linguagem de primeira ordem para uma *lógica da Indistinguibilidade*, que chamaremos abreviadamente L. Nosso sistema é baseado na lógica de Schrödinger, proposta por da Costa (da Costa [2008]) e generalizado para linguagens de ordem superior, como em da Costa e Krause [1994], [1997], e no sistema de lógica da Indiscernibilidade apresentado por Krause [2007] cap. 3. Utilizaremos os seguintes símbolos primitivos:

(a) Conectivos: $\rightarrow$ (implicação) e $\neg$ (negação);
(b) Quantificador universal: $\forall$ (para todo);
(c) Pontuação:), (, , (parênteses e vírgula);
(d) Uma coleção enumerável de variáveis individuais de primeira espécie $x_1, x_2, \ldots, x_n, \ldots$, e uma coleção qualquer de constantes individuais de primeira espécie $a_1, a_2, \ldots, a_n, \ldots$; claro que este modo de falar pode ser tornado adequadamente preciso e independente de noções como 'enumerável'. Por exemplo, para variáveis de primeira espécie poderíamos usar os

símbolos *x* e |, de forma que as variáveis de primeira espécie seriam expressões (sequencias finitas de símbolos) da forma *x*, *x*|, *x*||, etc. O mesmo pode ser dito das demais situações similares aqui apresentadas.

(e) Uma coleção enumerável de variáveis individuais de segunda espécie $X_1, X_2, \ldots, X_n, \ldots$, e uma coleção qualquer de constantes individuais de segunda espécie $A_1, A_2, \ldots, A_n, \ldots$;

(f) O símbolo de predicado binário '=' para a identidade e o símbolo de predicado binário '≡' para indistinguibilidade;

(g) Para cada número natural $n > 0$, uma coleção eventualmente vazia de símbolos de predicados de peso n.

Os outros conectivos, disjunção, conjunção e bi-implicação podem ser definidos da maneira usual, assim como o quantificador existencial. Um *termo* é uma variável ou uma constante individual. Os termos podem ser divididos, de modo evidente, em termos de primeira e segunda espécie. Por brevidade, usaremos as letras t_1, t_2, t_3, etc., como meta-variáveis para termos de qualquer das duas espécies, e x, y e z sem índices como meta-variáveis para variáveis de qualquer das duas espécies. Outra convenção que passaremos a utilizar será denotar por 'm-termos' os termos de primeira espécie e 'M-termos' os termos de segunda espécie.

Intuitivamente, os m-termos representarão as entidades básicas da microfísica, tal como descritas por alguma versão da mecânica quântica não-relativística, e os M-termos representarão os objetos macroscópicos. O objetivo, como comentamos acima, é fazer com que a identidade se aplique apenas a M-termos, e não a m-termos, pois para estes não faria sentido falar em identidade ou em diversidade, e que a indistinguibilidade se aplique a todos os objetos, desde que sejam da mesma espécie. Formalmente, isso se obtém ao se impedir, na definição de fórmula, que $t_i = t_j$ seja bem formada caso t_i ou t_j sejam m-termos, e ao impormos que t_i e t_j sejam da mesma espécie para que $t_i \equiv t_j$ seja bem formada. Com exceção destas restrições, a definição de fórmula também segue a usual, e como se pode notar a partir dos postulados que daremos, a lógica clássica se aplica de maneira usual aos M-termos. Mais especificamente, temos:

Definição [Fórmulas atômicas] Se t_i e t_j são termos da mesma espécie, então $t_i \equiv t_j$ é fórmula atômica. Se t_i e t_j são termos de segunda espécie, então $t_i = t_j$ é fórmula atômica. Se P é um símbolo de predicados de peso n, outro que a identidade ou a indistinguibilidade, e $t_1 \ldots t_n$ são n termos, então $Pt_1 \ldots t_n$ é fórmula atômica.

Definição [Fórmulas] As fórmulas de L são: (i) As fórmulas atômicas; (ii) Se α é fórmula, $\neg\alpha$ é fórmula; (iii) Se α e β são fórmulas, então $\alpha \rightarrow \beta$ é fórmula; (iv) Se α é uma fórmula e x é uma variável, então $\forall x\alpha$ é uma fórmula; (v) Apenas são fórmulas as expressões dadas pelas cláusulas anteriores.

É importante enfatizar: $t_i = t_j$ será fórmula apenas se ti e tj forem ambos termos de segunda espécie. Nossa definição de fórmula proíbe que expressões como, por exemplo, $x_1 = X_1$, ou $A_2 = x_1$ ou ainda $A_1 = a_2$ sejam bem formadas. No entanto, como dissemos anteriormente, a relação de indistinguibilidade vai se manter entre os dois tipos de objetos. Teremos que $t_i \equiv t_j$ sempre será uma fórmula bem formada, mas devemos observar a restrição de que t_i e t_j sejam de mesma espécie, ou seja, a definição de fórmula proíbe que, por exemplo, $x_1 \equiv X_1$, $A_2 \equiv x_1$ e $A_1 \equiv a_2$ sejam fórmulas.

O seguinte conjunto de postulados pode ser utilizado para L (outra formulação de uma lógica de primeira ordem com uma relação de indistinguibilidade pode ser encontrada nos postulados para a lógica da Indiscernibilidade, proposta em Krause [2007] cap. 3):

1. $\alpha \rightarrow (\beta \rightarrow \alpha)$
2. $\alpha \rightarrow (\beta \rightarrow \gamma) \rightarrow ((\alpha \rightarrow \beta) \rightarrow (\alpha \rightarrow \gamma))$
3. $((\neg\alpha \rightarrow \neg\beta) \rightarrow ((\neg\alpha \rightarrow \beta) \rightarrow \alpha))$
4. $\alpha \rightarrow \beta, \alpha / \beta$ (MP)
5. $\forall x\alpha(x) \rightarrow \alpha(t)$, com x e t da mesma espécie, e t livre para x em $\alpha(x)$.
6. $\beta \rightarrow \alpha(x) / \beta \rightarrow \forall x\alpha(x)$, onde x não ocorre livre em β.
7. $t = v \rightarrow (\alpha(t) \rightarrow \alpha(v))$, com t e v termos de segunda espécie, além das restrições usuais.
8. $\forall x(x \equiv x)$.
9. $\forall x \forall y(x \equiv y \rightarrow y \equiv x)$.
10. $\forall x \forall y \forall z(x \equiv y \wedge y \equiv z \rightarrow x \equiv z)$.
11. $\forall X \forall Y(X \equiv Y \rightarrow X = Y)$.

É importante perceber que os postulados que envolvem a relação de identidade estão formulados para termos de segunda espécie, principalmente o postulado 11, que intuitivamente significa que se dois macro-objetos são indiscerníveis, então eles são idênticos. Em breve veremos que a recíproca é teorema de L.

Conceitos sintáticos como os de demonstração, dedução a partir de um conjunto de fórmulas, teorema, ocorrências livres e ligadas de variáveis entre outros também são definidos da maneira usual. O Teorema da Dedução também pode ser demonstrado da maneira usual.

Temos agora o prometido teorema de que, para M-objetos, a igualdade implica a indistinguibilidade.

Teorema. $X=Y \rightarrow X\equiv Y$

Demonstração:

1. $X=Y$ (hipótese)
2. $X=Y \rightarrow (X\equiv X \rightarrow X\equiv Y)$ (postulado 7)
3. $(X\equiv X \rightarrow X\equiv Y)$ (1,2 Modus Ponens)
4. $\forall X(X\equiv X)$ (postulado 8)
5. $\forall X(X\equiv X) \rightarrow X\equiv X$ (postulado 5)
6. $X\equiv X$ (4, 5 Modus Ponens)
7. $X\equiv Y$ (3, 6 Modus Ponens)
8. $X=Y \rightarrow X\equiv Y$ (1-7 Teorema da Dedução)

Assim, com este teorema e o postulado 11, temos que para M-objetos, $X=Y \leftrightarrow X\equiv Y$, conforme prometemos anteriormente. No entanto, para m-objetos, este bicondicional não pode ser demonstrado. Ainda, neste caso, a relação de indistinguibilidade é caracterizada apenas como uma relação de equivalência, sem valer necessariamente o esquema da substituição dado pelo postulado 7. Isto a caracteriza sintaticamente como uma relação mais fraca do que a identidade. No entanto, como ainda não especificamos como devemos interpretar estes símbolos, nada impede, certamente, que ao se fazer uma semântica para L se interprete a relação de indistinguibilidade também como a relação de identidade, que satisfaz os axiomas de L dados para o símbolo de indistinguibilidade.

Também temos como teorema de L a reflexividade da identidade, que não precisa ser postulada. Com isto, a relação de identidade possui em L as duas propriedades que são usualmente utilizadas como

axiomas para a relação de identidade na lógica clássica de primeira ordem: reflexividade e substituição. Com estas propriedades, como se sabe, é possível demonstrar, por exemplo, que a relação de identidade é simétrica e transitiva.

Teorema. X = X

Demonstração:

1. $\forall X(X \equiv X)$ (postulado 8)
2. $\forall X(X \equiv X) \rightarrow (X \equiv X)$ (postulado 5)
3. $X \equiv X$ (1,2 Modus Ponens)
4. $\forall X(X \equiv X \rightarrow X = X)$ (postulado 11)
5. $\forall X(X \equiv X \rightarrow X = X) \rightarrow (X \equiv X \rightarrow X = X)$ (postulado 5)
6. $(X \equiv X \rightarrow X = X)$ (4,5 Modus Ponens)
7. $X = X$ (3, 6 Modus Ponens)

Não faremos a demonstração da simetria e transitividade da identidade para M-objetos aqui, pois o procedimento é o mesmo que na lógica clássica. Uma demonstração alternativa para estes fatos pode ser fornecida utilizando-se a equivalência para M-objetos entre identidade e indistinguibilidade e os postulados 8 e 9.

É interessante notar também que a lógica clássica de primeira ordem está de certo modo 'contida' na lógica L. Isto ocorre pelo fato de que, intuitivamente, os postulados de L, quando restritos aos M-termos, podem ser tomados como um conjunto de postulados para a lógica clássica. Falando mais rigorosamente, é possível se estabelecer uma tradução da lógica clássica em L, mostrando que os postulados da lógica clássica, quando traduzidos em L, são teoremas de L.

Sem dificuldade, podemos estender nossa lógica a uma lógica de ordem superior (teoria simples de tipos) e a uma teoria de conjuntos, que será em muito semelhante à teoria de quase-conjuntos *Q* (French e Krause [2006], cap.7).

3 Semântica clássica para L e seus problemas

Ao apresentar seu sistema de lógica, da Costa discutia uma interpretação pretendida, que deveria ser erigida de modo que as intuições básicas que serviram de base para se formular esta lógica fossem pre-

servadas. As variáveis individuais de segunda espécie percorreriam um conjunto no sentido usual, e as constantes individuais de segunda espécie nomeariam elementos deste conjunto. Por outro lado, as variáveis de primeira espécie deveriam percorrer uma coleção de não-indivíduos, e as constantes de primeira espécie deveriam nomear tais elementos. Os símbolos de relação de peso n, como usual, denotariam coleções de n-uplas de elementos destas coleções, e o símbolo de relação de indistinguibilidade, em nosso caso, deve denotar uma relação que simule a indistinguibilidade.

Este procedimento, no entanto, se conduzido da maneira usual, utilizando uma teoria de conjuntos ao estilo de ZFC como metalinguagem, suscita vários problemas filosóficos, caso queiramos preservar as intuições que deram origem à lógica de Schrödinger e a lógica da Indistinguibilidade, pois parece inviabilizar nosso comprometimento com uma ontologia de não-indivíduos relativamente aos objetos denotados pelos termos de primeira espécie. Isto ocorre, entre outros motivos, porque, como veremos com mais detalhe abaixo, nas teorias de conjuntos usuais, nas quais usualmente fundamentamos a semântica para linguagens formais como as que estamos discutindo, a identidade sempre faz sentido para todos os elementos do conjunto, e assim acabamos re-introduzindo a identidade para estes objetos através da metalinguagem.

Para discutirmos com mais rigor estes problemas, apresentamos a partir de agora um esboço de uma semântica clássica para L, onde o termo 'clássica' refere-se ao fato de que esta semântica é formulada na teoria de conjuntos ZFC. Como este modo de proceder é bastante conhecido, não seremos rigorosos e faremos apenas o suficiente para que possamos apresentar adiante, com mais detalhes, os problemas ocasionados por esta semântica quando desejamos tratar alguns elementos do domínio como não-indivíduos. Devemos enfatizar que as dificuldades aqui apresentadas não são uma exclusividade da lógica L e da lógica de Schrödinger, mas sim uma dificuldade que qualquer sistema de lógica para-reflexiva deve enfrentar. Em geral, como estes sistemas violam alguma forma do princípio de identidade, a semântica formal feita para eles deveria ser tal que os objetos do domínio da estrutura na qual interpretamos a linguagem preservassem esta característica. No entanto, como veremos no caso particular que estamos tratando, dificuldades surgem quando desejamos estabelecer

uma semântica para estes sistemas utilizando uma teoria de conjuntos clássica (ver também as discussões em da Costa e Bueno [2009]).

Procedendo da maneira conhecida, e seguindo as sugestões de da Costa ([2008] pp. 140, 141), queremos interpretar os símbolos não-lógicos da linguagem de L em uma estrutura **e** = <D, I>, onde:

1. D é um conjunto não-vazio que vai fazer o papel de domínio da estrutura. No presente caso, impomos ainda que $D = D_1 \cup D_2$, com $D_1 \cap D_2 = \varnothing$. Assumimos que as variáveis de primeira espécie tomam valores em D_1, e as de segunda em D_2.
2. Quanto a I, ela será a função denotação, e atribui às relações de identidade e indistinguibilidade e aos símbolos não-lógicos da linguagem elementos de D da seguinte forma:

2.1. Aos símbolos de predicados de peso n da linguagem de L, outros que a identidade e a indistinguibilidade, a função I associa da maneira usual um subconjunto de D^n.

2.2. Temos que: $I(a_i) \in D_1$, ou seja, às constantes de primeira espécie associam-se elementos de D_1, e $I(A_i) \in D_2$, ou seja, às constantes de segunda espécie associam-se elementos de D_2.

2.3. Ao símbolo de identidade associamos o conjunto $\{<x,y>: x,y \in D_2 \text{ e } x=y\}$.

2.4. Ao símbolo de indistinguibilidade ≡ atribuímos uma relação R em D tal que R é relação de equivalência.

Com a semântica acima esboçada, é possível obter resultados como a correção e completude de L com relação a esta semântica, da maneira usual, com as convenientes adaptações para a linguagem bissortida (Mendelson [1987] Cap. 2). No entanto, do ponto de vista filosófico, há uma série de problemas com esta semântica, problemas esses que surgem quando interpretamos os termos de primeira espécie como denotando entidades quânticas, tendo-se em vista que queremos sustentar, seguindo nossa leitura de Schrödinger e de outros autores, que elas são certo tipo de não-indivíduos. Alguns destes problemas já foram apontados por da Costa (da Costa [2008] pp. 140, 141), e outros ainda podem ser encontrados em Krause [2002].

Passamos agora a apresentar alguns destes problemas. Com eles, desejamos sugerir que é relevante buscar-se uma semântica mais ade-

quada para L do ponto de vista de suas motivações, ou seja, uma semântica na qual estes problemas possam ser superados. Uma das alternativas propostas na literatura é que se empregue como metalinguagem a Teoria de Quase-Conjuntos, uma teoria que permite formar coleções de objetos indistinguíveis mas não idênticos, mas não entraremos nestes detalhes neste trabalho (ver French e Krause [2006], cap. 7 e 8).

Do ponto de vista das motivações que deram origem à lógica da indistinguibilidade, o primeiro problema com essa semântica clássica começa com a escolha do domínio: nosso conjunto D_1 não poderia ser um *conjunto* no sentido comum das teorias de conjuntos usuais, em particular, não poderia ser um conjunto de ZF, a teoria de conjuntos que se utiliza usualmente como metalinguagem. Isto já havia sido apontado por da Costa ([2008], p. 140), e se deve ao fato de que essas teorias estão comprometidas com uma *noção cantoriana* de conjunto, no sentido de que conjuntos são coleções de objetos distintos uns dos outros, o que pressupõe a validade irrestrita da teoria da identidade para estes objetos e para seus elementos. Isto pode ser visto também como resultando do fato de que a lógica subjacente à teoria de conjuntos ZFC e todas as outras teorias de conjuntos clássicas é a lógica clássica, na qual a identidade se aplica sem restrições a todos os objetos, impedindo que para alguns deles, que deveriam representar os não-indivíduos, a identidade ou diferença não se aplique.

Ainda, é importante perceber que de acordo com o princípio de extensionalidade, dois conjuntos são idênticos se e somente se tiverem *os mesmos* elementos, o que depende, como se vê, de um conceito sensato de identidade. Além disso, os axiomas da teoria de conjuntos utilizada implicam que sempre é possível formar o conjunto unitário de um elemento dado, que será diferente do conjunto unitário de qualquer outro elemento (distinto do primeiro) por extensionalidade. Ou seja, um 'conjunto clássico' é um conjunto de *indivíduos* (no sentido por nós definido anteriormente), distinguíveis uns dos outros, para os quais a identidade sempre faz sentido. Para expressar este ponto com mais rigor, pode-se dizer que $<V, \in>$, o 'modelo' pretendido para ZF, se visto como uma estrutura matemática,[7]

[7] Este ponto, no entanto, é sutil. Não podemos elaborar um modelo de ZF na própria ZF (suposta consistente), como atesta o segundo teorema de incomple-

é uma estrutura rígida, mas não entraremos em detalhes aqui (ver Krause e Coelho [2005]).

Caso adotemos em nosso domínio o conjunto D_1 como um 'conjunto clássico', estaremos claramente reintroduzindo a identidade para estes elementos via metamatemática, ainda que queiramos impedir esse fato em nossa linguagem objeto. Desta forma, como salientou da Costa, estaremos abandonando a motivação inicial segundo a qual a identidade não deveria fazer sentido para certas entidades (que seriam os elementos de D_1). É importante enfatizar isto: uma das motivações para se propor L era possibilitar que se tratasse sensatamente com não-indivíduos, e se esta motivação não for observada na metalinguagem, aparentemente, não se terá restringido nada com as mudanças sintáticas da linguagem de L (para discussões ainda mais gerais sobre este tópico, ver da Costa, Bueno e Béziau [1995]).[8]

Outro aspecto no qual a maneira usual de fazer semântica entra em conflito com as motivações de L diz respeito à interpretação das constantes. Se quisermos que a identidade não tenha sentido para os elementos de D_1, não é possível que a função interpretação atribua a cada constante de primeira espécie um único e bem determinado elemento de D_1. Isto ocorre porque se os elementos de D_1 forem imaginados como denotando os *quanta*, de acordo com a interpretação que estamos supondo, não faz sentido nomeá-los desta maneira, pois se forem indistinguíveis, a princípio não podemos identificá-los nem distingui-los. Explicando um pouco mais este ponto, o problema é que podemos dar um 'nome' para uma partícula, por exemplo, cha-

tude de Gödel. Os modelos de ZF, caso existam, devem ser buscados em teorias mais fortes. Para certos conjuntos de axiomas, no entanto, podemos encontrar 'modelos internos' (no sentido de Gödel), mas não discutiremos este ponto aqui.

[8] Este tipo de situação ocorre freqüentemente quando se propõe um sistema de lógica que viole alguma das chamadas leis da lógica usual. Em geral, a metalinguagem utilizada para se fazer a semântica para estes sistemas pode ser considerada como sendo ZF, que pressupõe a validade da lei que se pretende derrogar, e acaba por nos comprometer com esta lei na metalinguagem. Por exemplo, na lógica intuicionista, desejamos entre outras coisas, que a lei do terceiro excluído não tenha validade geral, mas, se fizermos semântica para esta lógica em ZF, estaremos nos comprometendo, na metalinguagem, com a validade irrestrita desta lei, o que não é intuicionisticamente aceitável. Assim, semânticas distintas, que sejam aceitáveis de um ponto de vista intuicionista devem ser buscadas.

mando Eddy (nome bastante usado na literatura, bem como Priscilla e Astrid) a um elétron aprisionado em um aparato laboratorial. No entanto, esse 'nome' não resiste a uma eventual permutação de partículas, por exemplo, desfazendo e refazendo o experimento.[9] Neste caso a questão sobre se o Eddy preso no segundo experimento é ou não o mesmo que estava aprisionado no aparato quando fizemos o experimento pela primeira vez simplesmente não faz sentido, pois neste contexto os nomes não podem operar como designadores rígidos. (Isso também é salientado em Dalla Chiara1985; ver French e Krause op.cit., p.225).

Insistindo um pouco neste ponto, o problema não é simplesmente que se fizermos a suposição de que seja possível atribuir nomes no sentido usual para os quanta, haveria um problema em saber qual, dentre vários elementos indistinguíveis, é o portador de determinado nome, pois isso denotaria unicamente uma limitação epistemológica. O problema, no caso dos não-indivíduos, é ainda mais sério do que este; é que para eles, segundo a interpretação que estamos adotando, esta questão deixa de ser significativa, pois quando seguimos a interpretação acima apontada de Schrödinger, não faz sentido perguntar nem mesmo qual dentre várias entidades indistinguíveis recebeu o nome, e então a questão torna-se um problema de indeterminação ontológica, e insistir na rotulação através de nomes seria desistir da suposição de que estas entidades são realmente não-indivíduos segundo a nossa caracterização.

O problema com os nomes, mencionado acima, gera ainda outro grande inconveniente quando se pretende dar seqüência às definições anteriores. Usualmente, dois caminhos se apresentam. Por um lado, podemos definir a verdade para sentenças através do uso da linguagem diagrama relativamente ao domínio de interpretação, ampliando-se a linguagem ao acrescentar novos nomes, um para cada elemento do domínio, e então fornecer as cláusulas usuais por indução ou, por outro lado, podemos fazer esta definição através da noção de satisfatibilidade de uma fórmula por seqüências infinitas de elementos do domínio. Ambas alternativas são problemáticas do ponto de vista de nossos objetivos.

Com relação ao primeiro caminho apresentado, se não podemos

[9] Para mais discussões sobre este ponto, ver Krause [2006].

nomear os elementos de D_1, como já apontamos acima, também não é possível ampliar a linguagem de L com constantes individuais que sejam cada uma delas um nome para um elemento determinado do domínio, pois em particular, pelo mesmo argumento acima, não faz sentido dar um nome para cada elemento de D_1. Deste modo, caso decidamos nos manter fielmente de acordo com as motivações de L, caso adotemos uma postura 'schrödingeriana' com respeito às partículas, que por motivos de argumentação supomos que pertencem a D_1, então nos parece não ser possível formar a linguagem diagrama de L.

A situação descrita no parágrafo anterior não melhora caso se opte pelo segundo caminho apresentado acima, qual seja, definir a verdade para sentenças através da relação de satisfatibilidade de uma fórmula por seqüências de objetos do domínio. Na verdade, seria possível, seguindo-se este caminho, ter uma atitude ainda mais radical, optando-se por abandonar completamente qualquer referência a nomes, eliminando-se da linguagem todas as constantes de primeira espécie (e também de segunda), tendo-se em vista alguns dos argumentos acima. No entanto, como dissemos, caso resolva-se prosseguir desta maneira para definir a relação de satisfatibilidade, utilizando-se seqüências infinitas de objetos do domínio, tal como pode ser feito no caso usual (ver Mendelson [1987] cap. 2), os problemas com relação às motivações subjacentes a L não desaparecem.

Neste caso, o problema é que não se podem formar as seqüências de objetos sem especificar quais são os objetos, sem rotulá-los, uma vez que devemos ter claro em qualquer seqüência, para fins da definição de satisfação, qual é o seu n-ésimo elemento para qualquer n, ou seja, é preciso identificá-lo rotulando-o, e novamente voltamos ao problema mencionado acima de que os não-indivíduos não são rotuláveis deste modo. Ainda, no mesmo sentido, como uma seqüência é uma função do conjunto dos naturais no conjunto domínio, é preciso que para cada n natural se atribua um único elemento de D, em particular, de D_1, e isto significa que para cada natural esteja bem especificado o elemento de D que lhe corresponde, dadas as características de uma função em ZF, e que tenhamos critérios, novamente, de identificação para os elementos de D_1.

Um terceiro inconveniente com o qual nos deparamos diz respeito à atribuição de símbolos de predicados a subconjuntos de D,

ou seja, à especificação de extensões dos predicados da linguagem que se está interpretando. Novamente trata-se de conflito entre as motivações para a lógica L e a maneira usual de se fazer semântica. O problema é que não podemos dizer que certo predicado que verse sobre entidades quânticas determina um único conjunto com, digamos, n elementos. Na semântica usual todos os predicados são precisos ("sharp", para contrastá-los com predicados vagos), no sentido de que atribuímos a cada símbolo de predicados da linguagem um subconjunto de D^n (a extensão do predicado), de modo que qualquer n-upla de elementos do domínio pertence à extensão de um predicado n-ário ou não pertence (dado que o princípio do terceiro excluído vale na semântica calcada na lógica usual). No entanto, "...na física quântica, existem certos predicados que são *sharp* no sentido de que os físicos sabem muito bem quais condições um indivíduo deve obedecer para ter a propriedade associada pelo predicado, mas [aparentemente] existem objetos vagos, que nos induzem a considerar um tipo de 'ignorância ontológica' neste caso. Isto mostra que a relação entre os predicados (que estão pelas *intensões* de certos conceitos) e as suas correspondentes *extensões* (o conjunto dos indivíduos que tem a propriedade atribuída pelo predicado) se torna distinta da semântica standard." (Krause [2002] p. 78).

Assim, a situação da extensão de predicados, no caso da semântica que visa tratar de objetos da física quântica, é tal que não podemos determinar um conjunto bem definido de n-uplas para cada símbolo de predicados. Ainda temos, como no caso clássico que, dada uma n-upla de elementos do domínio, ela satisfaz ou não um predicado deste tipo. No entanto, para qualquer n-upla que pertence à extensão do predicado, outras n-uplas, indistinguíveis da primeira mas que podem não pertencer à coleção que determina a extensão também satisfarão o predicado. A satisfação de um predicado por uma n-upla depende mais do tipo de objetos que compõe a n-upla do que do fato de eles pertencerem ou não à extensão do predicado, pois todas as n-uplas de um certo tipo satisfarão um predicado, caso alguma delas satisfaça, ou seja, se uma n-upla de elementos do domínio satisfaz o predicado, então, qualquer n-upla indistinguível dela também satisfará.

4 Conclusão

Tendo em vista problemas como os mencionados acima, que se originam quando tentamos estabelecer uma semântica clássica para L de forma a manter as intuições que deram origem a esta lógica, segundo a qual os objetos com os quais trata a mecânica quântica não-relativista são certo tipo de não-indivíduos, aparentemente duas alternativas se sugerem, e que nos possibilitam superar estas dificuldades: abandonar o comprometimento com a ontologia de não-indivíduos e aceitar as dificuldades que surgem quando se adota esta posição relativamente à mecânica quântica, ou mudar a metalinguagem na qual se estabelece a semântica para linguagens como a de L. Newton da Costa sugeriu que, tendo em vista estas dificuldades, se criasse uma teoria de Quase-Conjuntos (da Costa [2008] p. 140), na qual fosse possível tratar de coleções de objetos indistinguíveis. Esta teoria pode ser vista em French e Krause [2006], cap. 7. Nela é possível erigir uma semântica para L que permite manter suas motivações metafísicas, mas a apresentação desta teoria e sua discussão estão fora do escopo deste trabalho. Este modo de proceder, apesar de nos permitir superar algumas das dificuldades apresentadas aqui e dar rigor aos termos nos quais uma ontologia de não-indivíduos pode ser formulada, ainda assim não encerra o debate filosófico sobre qual ontologia é a mais adequada para a Mecânica Quântica. No entanto, acreditamos que sistemas de lógica como o proposto, e outros, como a lógica de Schrödinger e a teoria de Quase-Conjuntos, contribuem para que a opção por uma ontologia de não-indivíduos se torne cada vez mais razoável, e para que formalismos que estejam mais de acordo com esta ontologia sejam erigidos para contribuir na discussão metafísica.

J.R. Arenhart
Universidade Federal da Fronteira Sul

D. Krause
Universidade Federal de Santa Catarina

Referências

da Costa, N.C.A. 1979. *Ensaio sobre os fundamentos da lógica*, São Paulo, Hucitec/ EdUSP 3ª. Ed. 2008.

da Costa, N.C.A. e Bueno, O. 2009. Lógicas não-reflexivas. In *Revista Brasileira de Filosofia*, ano 58, volume 232, 181-196.

da Costa, N.C.A. e Bueno, O., Béziau, J-Y. 1995. What is semantics? A brief note on a huge question. In *Sorites – Electronic Quarterly of Analytical Philosophy* 3, p. 43-47.

da Costa, N.C.A. e Krause, D. 1994. Schrödinger logics. In *Studia Logica* **53** (4), 1994, 533-550.

da Costa, N.C.A. e Krause, D. 1997. An Intensional Schrödinger Logic. In *Notre Dame Journal of Formal Logic* **38** (2), 1997, pp. 179-194.

Dalla Chiara, M.L. 1985. Names and descriptions in quantum logic. In *Recent Developments in Quantum Logics*, ed. by P. Mittelstaedt and E.W. Stachow. Manheim: Bibliografishes Institut, 189-202.

French, S. e Krause, D. 2006. *Identity in Physics. A historical, philosophical and formal analysis*. Oxford: Oxford University Press.

Krause, D. 2002. Why quasi-sets? In *Boletim da Sociedade Paranaense de Matemática* **v. 20** 1/2, 2002, 73-92.

Krause, D. 2006. Is Priscilla, the trapped positron, an individual? Quantum physics, the use of names, and individuation. A aparecer em *Arbor* (Spain).

Krause, D. 2007. *La metafísica de los quanta*. Manuscrito em preparo.

Krause, D., Coelho, A.M.N. 2005. Identity, Indiscernibility and Philosophical Claims. In *Axiomathes* **15** (2) pp. 191-210.

Mendelson, E. 1987. *Introduction to Mathematical Logic*. 3ª ed., Wadsworth & Brooks.

Muller, F.A. e Saunders, S. 2008. Discerning Fermions. In *The British Journal for the Philosophy of Science*, 59(3) pp. 499-548.

Muller, F.A. e Seevinck, M.P. 2009. Discerning Elementary Particles. A aparecer em *Philosophy of Science*.

Saunders, S. 2006. Are quantum particles objects? In *Analysis*, 66.1, January, pp. 52-63.

Schrödinger, E. 1952. *Science and Humanism*. Cambridge: Cambridge Un. Press.

What Is Existence?[1]

João Branquinho
University of Lisbon and LanCog Group

BIBLID [0873-626X (2012) 34; pp. 575-590]

1 Introduction

This paper has a negative and a positive claim. The negative claim is that the Frege-Russell account of existence as a higher-order predicate is mistaken and should be abandoned, even with respect to general statements of existence such as "Flying mammals exist" (where statements of this sort are supposed to be best accommodated by the account). The Frege-Russell view seems to be supported by two ideas. First, the idea that existence is entirely expressed by the existential quantifier of standard predicate logic. Second, the idea that the existential quantifier is a higher-order predicate, a predicate of predicates, not of individuals. I think that both ideas are wrong but will focus on the latter. By construing prima facie first-order statements such as "Flying Mammals exist" as higher-order predications such as "The Fregean Concept *Flying Mammal* maps at least one individual onto the True", the Frege-Russell view commits one - merely on the basis of the meaning it assigns to the existence predicate – to abstract objects such as concepts (Gottlob Frege), or propositional functions (Bertrand Russell), or classes (Rudolf Carnap), or properties, kinds, and so on. This cannot be right, I think.

The positive claim of the present paper is that, at least in the context of first-order discourse, the existence predicate is just what it seems to be: a bona fide first-order predicate (pace Kant, Hume, Frege, Russell and others). Three important ideas about existence are shared with the Frege-Russell conception of existence, though.

[1] I have been strongly influenced in this paper by the views advanced by Nathan Salmon, mainly in Salmon 1987 and Salmon 1998. The views endorsed here are similar to his views, but my motivation is different.

(1) Being and existence are one and the same thing: there is no difference between "Unicorns are not", or "There are no unicorns", and "Unicorns do not exist", or "There exist no unicorns". (2) To be is to be the value of a bound variable, to belong to a domain of quantification (Willard Quine). (3) Anti-Meinongianism, the idea that there are no non-existent objects (Russell). However, we diverge from the Frege-Russell tradition with respect to the following claim. (4) The best concept of existence, in the sense of the one that is best understood and best enables us to formulate ontological disputes, is a purely logical first-level concept defined in terms of existential quantification and identity. First-order statements of existence and non-existence like "Flying Mammals exist" and "Unicorns do not exist" are accordingly taken at face value and analyzed in terms of a logical first-order predicate of existence, the predicate "is (identical to) something". Reasons are given to prefer this notion of existence to other first-order non-logical notions that have been proposed in the literature, notions characterized in terms of predicates such as "is in space-time", "is concrete", "is causally efficacious", "is actual", "is real", etc.

We will also reflect upon the concept of existence by studying the logical form of statements of existence and non-existence, statements such as

Flying mammals exist
Unicorns do not exist
Venus (the planet) exists
Vulcan (the planet) doesn't exist

We are particularly interested in the logical and semantic status of the existence predicate involved therein. We want to determine what existence predicate we should have at the level of logical form that would correspond to the grammatical predicate "exist(s)" at the surface level.

The issue about the logical form of existence statements is a vexed issue in contemporary philosophical semantics, an issue that is far from having received a satisfactory treatment. On the other hand, I think that the search for an adequate existence predicate can only be correctly carried out if we first provide answers to a salient set of

general questions about existence. In what follows I introduce three such questions and three theses I want to endorse in answering them, such theses shaping the subsequent adoption of an appropriate existence predicate.

Availing ourselves of an appropriate existence predicate is highly important for purposes of meta-ontology, for it allows us to describe ontological disagreements, disagreements about what exists, as they should be viewed (at least sometimes): genuine disagreements, not merely verbal or terminological ones.

2 Existence and quantification

Here is the first of our three questions concerning existence.

Question 1 - Existence and Quantification

Is there any relation between the concept of existence and the concept of quantification, especially existential quantification?

The answer to this question that I would like to favor is this.

Thesis 1: Existence is not entirely expressed by the existential quantifier $, but there is an important connection between the two concepts: the concept of existential quantification should be seen as playing a central role in a correct characterization of the concept of existence (details later)

Of course, several philosophers have rejected Thesis 1. On one side of the opposition is the Frege-Russell view, also famously endorsed by Quine (Quine 1980: 12-13), on which existence is fully represented by the existential quantifier. We will come back to the Frege-Russell view later on. On the other side of the opposition is Meinongianism, defined in general as the view that some objects do not exist.

Meinongianism comes in a variety of versions, including the original views of Russel in *Principles of Mathematics* (Russell 1903), Terence Parsons's views in *Non-Existent Objects* (Parsons 1980), and more recent versions developed by Richard Routley (Routley 1980) and Graham Priest (Priest 2005**)**, known as Noneism (David Lewis coined this term in Lewis 1990). However, all brands of Meinongianism have in common the rejection of any sort of explanatory link

between the concepts of existence and quantification.

On the latter side of the opposition to Thesis 1 is also the apparently anti-Meinongian position recently advanced by Kit Fine (Fine 2009). Fine develops a set of interesting considerations with a view to rejecting any account of existence in terms of quantification. However, we believe it is wrong to separate in limine, from the point of view of explanation, these concepts. After all, there seems to be a strong intuitive sense in which the existential quantifier carries existential force, has ontological import. We regard as implausible the reading of $\exists$ as a merely "particular" quantifier (Priest), deprived of the ontological role of introducing at least one object of a domain of quantification. We prefer a moderate view, on which the existence predicate is still a logical predicate, but one only partially defined in terms of existential quantification (Thesis 1).

3 Is existence a first-level concept?

We turn now to our second question about existence.

Question 2 - Is existence a (first-order) predicate?

This is the old question of whether existence is, or can be a "real" predicate, a predicate like the others, a predicate of familiar things, a predicate like "flies", "is a mammal", "is famous", etc.

There are two extreme positions concerning this question, which I label the Old School and the Very Old School. We want to endorse the Very Old School, but let us take the Old School first.

(a) The Old School

This is basically the Frege-Russel conception of existence (See Russell 1988: 211 and Frege 1950: 64-65). It consists in giving a negative answer to Question 2 on the basis of two premises.

> **Premise 1:** $\exists$ is a higher-order predicate, a predicate of predicates, never applicable to entities of level 0 or individuals.

Roughly speaking, individuals are those entities that, in spite of being able to belong to classes, to instantiate properties, to be members of species and kinds, to be subsumed by Fregean concepts, to

be arguments of Russellian propositional functions, and so on, are not themselves classes, properties, species, kinds, Fregean concepts, propositional functions, and so on.

> **Premise 2**: The already mentioned claim that the concept of existence is entirely expressed by ∃.

These two premises entail the following claim, a claim also endorsed (at least in its negative version) by Kant and Hume.

> **Conclusion**: Existence is invariably a higher-order predicate, never a predicate of individuals.

Before critically examining the Old School, let us introduce the Very Old one.

(b) The Very Old School

This position gives an affirmative answer to Question 2 and consists in the following thesis.

> **Thesis 2**: Existence is a first-order predicate.

(As we shall soon see, we must be careful here and take Thesis 2 as presupposing a restriction of the universe of discourse to individuals.)

The claim that existence is, or can be, a first-order predicate is endorsed in all varieties of Meinongianism. It is also endorsed on the already mentioned, non- Meinongian, account proposed by Fine. It is further endorsed on the present view, which is not Meinongian either (see below). It is therefore a mistake to think that rejecting the claim that existence is a higher-order predicate entails embracing Meinongianism.

As noted, Thesis 2 has to be subjected to the important qualification that, in the context of our discussion of Question 2, we are dealing only with first-order discourse, with statements about individuals. Thus, the following statements would presumably be excluded from our discussion, for they are higher-order (or so we assume for the sake of argument):

Wolf and dog inter-breed.
There are animal species on the verge of extinction.
Humility is rare, cowardice despicable.
The class of prime numbers is infinite.

In contrast, the following statements would presumably be admitted (or so we assume for the sake of argument):

The wolf is more aggressive than the dog.
The dog has warm blood.
There are flying mammals.
Humility is a virtue.

Now if the above qualification were not made, Thesis 2 would be promptly refuted on the basis of statements such as

Primary colors exist.
The Dodo bird no longer exists.

Indeed, the existence predicate is clearly second-order here.

It is crucial to note that, even under the restriction to a domain of individuals, existence is still a higher-order predicate on the Frege-Russell view. Let us check this by considering seemingly first-order statements such as

(1) Flying mammals exist.
(2) Unicorns do not exist.

The Frege-Russell analysis is carried out in two steps. First, in the light of the Frege-Russell claim that existence is fully expressed by the existential quantifier, such statements are analyzed as

(1)' Something is a flying mammal
(1)' $\exists$x Flying Mammal x
(2)' Nothing is a unicorn
(2)' $\neg\exists$x Unicorn x

Second, the latter statements are in turn paraphrased into second-order statements such as (these are only examples)

(1)" The class of flying mammals is not empty
(1)" The property of being a flying mammal is instantiated
(1)" The Fregean concept Flying Mammal maps at least one individual onto the True
(1)" The propositional function Flying Mammal is possible
(2)" The class of unicorns is empty
(2)" The property of being a unicorn has no instances
(2)" The Fregean concept Unicorn maps no individual onto the True
(2)" The propositional function Unicorn is impossible

We believe that the second step of the Frege-Russell analysis is profoundly mistaken, that the proposed paraphrase in terms of higher-order predications is wrong.

Here are four objections to the Frege-Russell view.

Objection 1: Expressive Power

The Frege-Russell account does not seem to have the means to express, in the language of the theory, some existence and non-existence claims to which it is manifestly committed. In particular, it does not seem to have the means to express the anti-Meinongian statement "Everything exists" or "There are no non-existent objects". It is hard to see how these statements could be analyzed in the Frege-Russell style, how the existential quantifier could here give way to an appropriate higher-order predicate.

Objection 2: Ontological Inflation

The Frege-Russell treatment of the existential quantifier as a higher-order predicate has immediate anti-nominalist consequences, or (if you prefer) immediate Platonist or Realist consequences, which cannot be right in my view. A true statement of existence like "Flying mammals exist" ontologically commits us not only to things that are mammals and fly (these are individuals and concrete items), but

also to abstract objects such as classes, Fregean concepts, properties, propositional functions, etc. And even true statements of non-existence, such as "Unicorns do not exist", ontologically commit us to the very same sort of abstract objects (although they do not commit us to unicorns).

Note that we might have good reasons to introduce abstract objects, even of all the types in question, into our best ontology. But not merely on those grounds, not merely on the basis of a proposal about the meaning and logical form of statements of existence and non-existence.

Objection 3: Slippery Slope

This is an argument in the style of Frank Ramsey (see Ramsey 1925).

If a true predication of existence like "Flying mammals exist" were to be paraphrased into something like "The Fregean concept *Flying Mammal* maps at least one individual onto the True", then nothing would prevent us from paraphrasing in the same way virtually any predication, including common predications such as "Mammals have warm blood" and "Rover is a dog". The result would be something like "The Fregean concept *Having Warm Blood* maps onto the True any individual mapped onto the True by the Fregean concept *Mammal*" and "The Fregean concept *Dog* maps the individual Rover onto the True".

The same would go for paraphrases in terms of classes, properties, propositional functions, and so on. Any prima facie first-order predication would turn out to be, at bottom, higher-order in nature. I take it that this is a highly implausible consequence of the Frege-Russell account of the existence predicate.

Objection 4: The Intuitive Criterion of Difference for Thoughts

This Fregean principle, as formulated by Gareth Evans (see Evans **1982**: 21), states that thoughts or contents **p** and **q** are distinct if it is rationally possible to take conflicting propositional attitudes towards them, say believing **p** while not believing **q** or disbelieving **q**, believing **p** while doubting **q**, etc. Now it seems perfectly possible for a rational subject to accept "Flying mammals exist" and "Unicorns do

not exist" but at the same time to be in doubt about, or even reject, their Fregean paraphrases "The Fregean concept *Flying Mammal* maps at least one individual onto the True" and "The Fregean concept *Unicorn* maps no individual onto the True". The subject might so proceed on the basis of strong nominalist convictions, or because she is just skeptical about entities such as Fregean concepts. And it does not matter at all if the subject is right or wrong in doing so. The same would go for paraphrases in terms of classes, properties, propositional functions, and so on.

4 Existence and being

We turn now to our third question about existence.

> **Question 3 - Being and Existence**
>
> What is the relation between being, in the sense of being something, being an object, and existing, or having existence? Does being transcend in any sense existence? Should we claim that something does not exist, that some objects do not exist? Or should we rather claim that everything exists, that every object exists?

On the most usual versions of Meinongianism, there are objects that do not exist: the realm of being, of what can be quantified over or referred to (roughly speaking), is broader than the realm of existence, of objects in space-time (roughly speaking). On other versions of Meinongianism, we have only the weaker claim that some objects do not exist (the so-called particular quantifier "some" having no ontological or existential import). This is the case of the original views of Meinong, since he posits objects that do not have any form of being, such as chimeras and impossible objects. And is also the case of the Noneist views of Routley and Priest (for the same reason).

Noneism has the advantage of keeping Meinongianism immune to what is often seen as a serious objection to the position, namely that the distinction it often makes between being and existence makes little sense. As Quine remarks (Quine 1980: 3; Quine 1969: 100), there is no discernible difference between statements such as "There are prime numbers" and statements such as "There exist prime num-

bers". David Lewis (Lewis 1990) and Peter van Inwagen (van Inwagen 2008) argue in the same direction.

But there is another serious objection to Meinongianism, and this one also applies to the Noneist variety. The objection is that Meinongianism obliterates a distinction that should be made in any case between genuine reference, e.g. "The American who lives upstairs" (where the description has a referential use), and merely apparent reference, e.g. "The average American" or (perhaps) "My shadow". Meinongianism also obliterates, in the same vein, a distinction that should be made in any case between genuine quantification, e.g. (perhaps) "There are prime numbers", and merely apparent quantification, e.g. (perhaps) "There are intolerable fluctuations in the stock market". Such distinctions are obliterated on the Meinongian view because this view seems to be committed to the idea that any term that appears to denote something actually denotes something, and that any expression that appears to quantify over something actually quantifies over something. We find this idea unacceptable as it goes against basic Russelian wisdom. So we endorse the following anti-Meinongian thesis with respect to Question 3:

Thesis 3: Everything exists, there are no non-existent objects

We introduce below further reasons for rejecting Meinongianism and accepting Thesis 3.

5 The existence predicate

We note now that the existence predicate we are looking for will have to conform to Thesis 3, which means that it has to be an existence predicate **E** that satisfies the following principle

(E) $\forall x Ex$

In other words, we need an existence predicate that is true of every object and false of no object. That is to say, we want the extension of **E** to be the entire domain of quantification.

On the other hand, by Thesis 2, **E** has to be a first-order predicate (assuming a universe of discourse containing only individuals). Also, by Thesis 1, **E** has to be a predicate partially definable in terms of

existential quantification.

Finally, having our initial methodological remarks in mind, our existence predicate **E** should be conceptually clear and apt to correctly describe a wide variety of ontological disputes, disputes about what there is or exists, as they at least sometimes are, viz. as substantive disputes.

Now, the existence predicate **E** we are looking for, one that satisfies the set of Theses 1,2,3 and meets the above methodological requirements, is simply the familiar predicate **_is something, _is identical to at least one object** (See Quine 1969: 97; also Kripke 2011: 55, Footnote 6 and Salmon 1987:20-2). (Of course, I assume that our language contains the identity predicate among its logical constants.)

$$Ex = (df)\ \exists y\ x{=}y$$

Let us check this. If one is dealing with first-order discourse and our domain is a domain of individuals, then our existence predicate will invariably be a first-order predicate, a predicate of individuals, vindicating thus Thesis 2. On the other hand, our existence predicate is not primitive, since it is defined in terms of quantification and identity, vindicating thus Thesis 1. Also, it is a purely logical predicate, as it is characterized in terms of logical concepts only. Finally, it is a predicate that is entirely in order from the point of view of conceptual clarity, at least to the extent that logical concepts are entirely in order from that point of view.

Notice that "Everything exists", in symbols $\forall xEx$ or $\forall x\exists y\ x{=}y$, is a logical truth and thus (in some sense) a trivial truth. Our existence predicate is a tautologous predicate and therefore also a trivial predicate (in some sense). However, such triviality can be somehow mitigated if we notice that ontological disputes are not automatically solved on that basis (Quine 1980:1). To exist, or to be, is to belong to a domain of quantification, and everything belongs to a domain of quantification, but that does not by itself tell us what to include in the domain of quantification, it does not by itself tell us what we should put among everything. We might still want or not want to include mere possibilia, fictional objects, chimeras and other intentional objects, universals, numbers, material objects, arbitrary fusions of ma-

terial objects, temporal parts, etc.

We go back to the Meinongian view now. What other choices would be available for a first-order existence predicate **E**? Here is a list of some of the usual proposals, most of them having a clear Meinongian motivation.

(a) Ex = x is causally efficacious (Priest)
(b) Ex = x is actual (in the modal sense)
(c) Ex = x is concrete
(c)' Ex = x is in space-time (Russell)
(d) Ex = x is real, where "real" is a primitive predicate (Fine 2009: 168-9)
(e) Ex = x is a non-intencional object (MGinn 2000: 15-51)

The main problem with the Meinongian proposals (a)-(c)', and also with the quasi-Meninongian proposal (e), is a problem of meta-ontological inadequacy. Indeed, the characterizations proposed for the existence predicate **E** have the undesirable feature of entailing a rejection from the outset of a certain range of ontological positions, which would thus be counted as conceptually false, i.e. false merely in virtue of the concept of existence employed. Here are examples of such positions: "Universals exist", "Mere possibilia exist", "Classes exist", "Numbers exist". It might be replied that on the most usual versions of Meinongianism we could still have truths like "There are universals", "There are mere possibilia", "There are classes", "There are numbers", etc. But, as noted, the problem with those views is that they rely on a distinction between being and existence that it is hard to make sense of.

So the Meinongian view underlying proposals (a)-(c)' has immediate nominalist implications. On the other side, as we have seen, the Frege-Russell view has immediate anti-nominalist implications. Both are thus wrong for the same kind of reason.

The problem with proposal (d) is that it is not completely clear what "real" means; or, to be more cautious, one should at least say that its meaning is less clear than the meaning of our existence predicate.

I finish with a few brief remarks on logical form. How statements of existence and non-existence of central kinds should be analyzed

on the present view?

With respect to statements of singular existence and non-existence, the answer is readily available.

Singular Existence: a exists
$Ea, \exists y\ a=y$

Singular Non-existence: a does not exist
$\neg Ea, \neg\exists y\ a=y$

With respect to statements of general existence and non-existence, we need to be more careful. Take the former ones, first.

General Existence: Fs exist

This is a more complicated case, but for reasons given below we go for

$\exists x\ (Fx \wedge \exists y x=y)$

A statement like "Ostriches are fast" is ambiguous between a universal quantification, "All ostriches are fast", an existential quantification, "Some ostriches are fast", and a generic, "Ostriches are typically fast". By analogy, a statement like "Flying mammals exist" admits two reading (excluding the generic reading for obvious reasons).

Reading 1: Every flying mammals exist
$\forall x(MVx \rightarrow \exists y\ x=y)$

Fine reads this way and objects that if the existence predicate is our tautologous predicate, then the statement "Flying mammals exist" would turn out to be trivially true, as it would be a logical truth. Yet, there are some doubts about this. If one adopts a free logic that restricts the rule of introduction of existential quantification in the familiar way, and there are independent reasons to do it, then it is not clear that the statement is a logical truth.

At any rate, another, more serious, objection to reading 1 is that a statement like "Unicorns exist" would turn out to be true – vacu-

ously true, assuming that the domain of quantification does not contain unicorns.

One way of replying to this objection would be to replace the usual quantifiers of classical first-order predicate logic with generalized quantifiers, being thus able to block such undesirable assignments of truth-value; but we will leave the issue at this point.

Reading 2: Some flying mammals exist
$\exists x\ (MVx \wedge \exists yx{=}y)$

We prefer this reading, which is clearly not a logical truth. The existence predicate is indeed in a sense tautological: nothing is added by it if the domain of quantification already contains at least one flying mammal. But that is what should be expected given the logical nature of our existence predicate.

Given the analysis proposed for general existence, general non-existence has a straightforward rendering.

General Non-existence: Fs do not exist
$\forall x(Fx \rightarrow \neg\exists yx{=}y)$

We close with an interesting observation. Take the symbolizations proposed for general existence and non-existence.

Fs exist
$\exists x\ (Fx \wedge \exists yx{=}y)$

Fs do not exist
$\forall x(Fx \rightarrow \neg\exists yx{=}y)$

It turns out that they are logically equivalent to the simpler symbolizations one finds in logic textbooks, namely

$\exists xFx$
$\neg\exists xFx$.

João Branquinho
University of Lisbon, LanCog Group

References

Evans, G. 1982. *The Varieties of Reference*, ed. by John McDowell. Oxford: Clarendon Press and New York: Oxford Unviversity Press.

Fine, K. 2009. The Question of Ontology. In *Metametaphysics. New Essays on the Foundations of Ontology*, ed. by David J. Chalmers, David Manley and Ryan Wasserman. Oxford: OUP, 157-77.

Frege, G. 1950. *The Foundations of Arithmetic. A Logico-Mathematical Inquiry into the Concept of Number.* Tr. by J.L. Austin. Oxford: Basil Blackwell.

Kaplan, D. 1979. Transworld heir lines. In *The Possible and the Actual*, ed. by M. Loux. Ithaca, NY: Cornell University Press, 88-109.

Kaplan, D. 1989a. Demonstratives. In *Themes from Kaplan,* ed. by Almog, J. Perry and H. Wettstein. Oxford: Oxford University Press, 481-563.

Kaplan, D. 1989b. Afterthoughts. In *Themes from Kaplan*, ed. by J. Almog, J. Perry and H. Wettstein. Oxford: Oxford University Press, 565-614.

Kripke, S. 1980. *Naming and Necessity*. Cambridge, MA: Harvard University Press.

Kripke, S. 2011. Vacuous Names and Fictional Entities. In *Philosophical Troubles.* S. Kripke, Collected Papers. Volume I. Oxford, Oxford University Press: 52-74.

Lewis, D.K. 1990. Noneism or Allism?, *Mind* 99: 23-31

McGinn, C. 2000. *Logical Properties. Identity, Existence, Predication, Necessary Truth*. Oxford: Clarendon Press.

Meinong, A. 1960. *On the Theory of Objects*. Translated by R. Chisholm, I. Levi and D. Terrell. In *Realism and the Background of Phenomenology*, ed. by R. Chisholm. Glencoe: The Free Press, 76-117.

Moore, G. E. 1936. Is Existence a Predicate? *Proceedings of the Aristotelian Society*, Supplementary Volume XV: 175-88. Reprinted in G.E. Moore, *Philosophical Papers*. London: George Allen and Unwin, 102-14.

Parsons, T. 1980. *Non-existent Objects*. New Haven, CT: Yale University Press.

Parsons, T. 1995. Are There Non-existent Objects. *American Philosophical Quarterly*, Vol. 19, No. 4, Oct., 1982 , 365-371

Priest, G. 2005. *Towards Non-Being. The Logic and Metaphysics of Intentionality.* Oxford, OUP.

Quine, W. V. O. 1969. Existence and Quantification. In *Ontological Relativity and Other Essays*. New York: Columbia University Press, 91-113.

Quine, W. V. O. 1980. On What There Is. In *From a Logical Point of View. Nine Logico-Philosophical Essays.* 2nd edition. Cambridge, Massachusetts and London, England: Harvard University Press.

Ramsey, F. P. 1925. Universals. *Mind*, New Series, Vol. 34, No. 136 (Oct., 1925), 401-417

Routley, R. 1980. *Exploring Meinong's Jungle and Beyond: An Investigation of Noneism and the Theory of Items.* Australian National University.

Russell, B. 1903. *Principles of Mathematics*. Cambridge, England, Cambridge University Press

Russell, B. 1905. On Denoting. *Mind* 14:479-93.

Russell, B. 1988. The Philosophy of Logical Atomism. In *The Collected Papers of Bertrand Russell,* Volume 8: The Philosophy of Logical Atomism and Other Essays, 1914-19. Ed. John Slater. London: Routledge, 155-244.

Salmon, N. 1987. Existence. In *Philosophical Perspectives*, 1: Metaphysics, ed. by James Tomberlin. Atascadero, Cal.: Ridgeview, 49-108. Reprinted in *Metaphysics, Mathematics and Meaning: Philosophical Papers*, Volume I. Oxford: Clarendon: 9-49.

Salmon, N. 1998. Nonexistence. *Nous* 32/3: 277-319. Reprinted in *Metaphysics,*

Mathematics and Meaning: Philosophical Papers, Volume I. Oxford: Clarendon: 50-90.
Van Inwagen, P. 1977. Creatures of Fiction. *American Philosophical Quarterly* 14/4 (1977): 299-308.
Van Inwagen, P. 2008. McGinn on Existence. *Philosophical Quarterly*, Volume 58 – Issue 230: 36-58
Williamson, T. 2013. *Modal Logic as Metaphysics*. Oxford: Oxford University Press.

Philosophy as a Protoscience

Cláudio F. Costa
Federal University of Rio Grande do Norte

BIBLID [0873-626X (2012) 34; pp. 591-608]

Where philosophy was, there science shall be.
Robert Nozick

My intention in this paper is to investigate the structural and dynamic relationships between philosophy and science, particularly the view that philosophy anticipates and leads into science. This investigation shades some light on the nature of both, philosophy and science, and on their mutual relations.

1 The philosophy in its Greek origins as a case study

When in search of an explanation for the nature of philosophy, a good starting point is to inquire as to its origins. As is commonly known, Occidental philosophy originated in Ancient Greece as a substitute for mythological and religious explanations. Instead of accepting explanations of the foundations and origins of reality based on the anthropomorphic projections of mythology, the early Greek philosophers realized that reality could also be explained speculatively, by appealing to impersonal (or nearly impersonal) *principles*, for example, water (Thales), air (Anaximenes), the infinite (Anaximander) and being (Parmenides), or living forces like love and hate (Empedocles)....[1] Questions that could help us to understand the na-

[1] In the history of philosophy, equivalents to these principles have continually been proposed: Plato's *ideas*, Aristotle's *substance*, Aquinas' *God*, Kant's *thing in itself*, Fichte's *I*, Hegel's *absolute*, Shopenhauer's *will*, Heidegger's *Being* and Wittgenstein's *unsayable*, played a similar foundational role.

ture of philosophy are in this case: What is the reason for this change in explanatory approach? What is the nature of this change?

A good reason for the shift from mythological to philosophical thought has been proposed by historians of philosophy such as W.K.C. Guthrie[2]. According to this author, Greek thinkers, having borrowed scientific knowledge (astronomical, physical, geometrical, arithmetic, etc.) from other cultures, were the first to consider such knowledge *in abstraction from practical applications*, namely, in the form of *theoretical generalizations*. We can see the best example of this attitude in Euclid's *Elements*, with its axiomatic-deductive method of proving theorems. It was this awareness of the explanatory power of theoretical generalization that presumably suggested to early Greek thinkers the possibility that questions once answered by means of the anthropomorphic metaphors of mythology and religion could instead be addressed in terms of abstract speculative generalizations, that is, in philosophical terms.

Although persuasive, this last explanation remains incomplete. Admittedly, the Greeks were the first to consider scientific generalization apart from its application. They were the first to axiomatize geometry, and they were able to produce physical and astronomical generalizations (such as, respectively, the measurement of specific gravity by Archimedes and the heliocentric hypothesis of Aristarchus). However, in order to explain the emergence of philosophical thought it is not enough to consider the emergence of explicit generalizations independently of their practical applications, for this is not a privilege of scientific explanation. Commonsensical explanation, for example, is also based on empirical generalizations, like those conveyed by sentences such as 'The sun always rises', 'Water quenches the thirst', 'Fire burns', etc., which are not scientific but have always been accepted as conveying obvious truths. Moreover, people were certainly always able to consider such trivial generalizations apart from practical concerns.

A more complete explanation for the emergence of philosophy in Greece seems to me the following. When they succeed in creating abstract scientific knowledge, Greek thinkers also achieved an intuitive understanding of the nature of the generalizations and ex-

[2] See Guthrie, W.K.C., *A History of Greek Philosophy,* vol. 1, p. 36 ff.

planations of science, as much for the formal sciences (geometric theorems) as for the empirical ones (physical and astronomical laws). Though at the very beginning they didn't have a philosophy of science (the first steps in this direction were taken later by Aristotle in his *Organon*), they certainly had an *idea* of the kind of hypothetical, predictive and explanatory procedures that are common to science in general – empirical as much as formal – what we could call an *idea of science*. Now, it seems that Greek philosophy was born from the speculative application of this idea of science to questions earlier approached exclusively by means of religion, like the question of the ultimate nature of the world and of our place in it. Provided with this new notional attitude, the first Greek philosophers attempted to proceed rationally, first by seeking to establish true generalizations based on some kinds of *data* (empirical or formal), and then by trying to explain some kinds of facts, whatever their nature, using these generalizations[3]. The first Greek philosophers pursued this aim by introducing vague principles (like water, air, infinitude, being) or forces (like heat and cold, love and hate), which might be interpreted as a first attempt to replace explanation relying on the actions and intentions of divinities with explanation based on the constitutive elements of reality and the impersonal laws regulating their transformations, often hovering midway between the two kinds of explanation.[4] It is not without reason that Thales, the first philosopher of the Occidental tradition, was also a scientist and a competent astronomer, who once accurately predicted a solar eclipse.

[3] A similar procedure applies even to philosophy understood as conceptual analysis: philosophers usually consider certain *data*, as they appear in examples, paradigmatic cases, thought-experiments, etc. in an effort to reach some kind of conceptual generalization, by means of which they attempt to explain a large set of conceptual applications.

[4] This phenomenon was already noted by Auguste Comte as he considered the passage from mythological to metaphysical thought. For a discussion, see C.F. Costa: *The Philosophical Inquiry: Towards a Global Account*, chap. 4.

2 Philosophy as a conjectural inquiry lacking consensual foundations

Assuming that Occidental philosophy arose from the speculative application of the idea of science to questions inherited from mythology and religion, how should we distinguish the activities of philosophers from those of scientists? – for, though there are some suggestions to the contrary,[5] there seems to be a considerable difference! The answer to this question brings us to what I regard as a central insight into the nature of philosophy. Even if philosophical activity resembles the general procedure of scientific inquiry, there is a fundamental difference in that philosophical explanation remains *merely conjectural* and, to this extent, *speculative*.[6]

But what do the words 'conjectural' and 'speculative' mean when we say that philosophical investigation remains conjectural or speculative? One answer is that an investigation is conjectural when it achieves only *hypothetical results*, and that this is the case when there is *no possibility of consensual agreement about the truth of their results*. Indeed, while in the sciences it is fairly easy to obtain consensual agreement on results, this kind of consensus is impossible in the muddy waters of philosophical inquiry. Consider the difference: The explanation of how levers work using the laws of levers, as conceived by the Greek scientist Archimedes, was one that everyone could verify practically and agree to. In contrast, the explanation of the generation and destruction of things through the action of the living forces

[5] See, for example, W.v.O. Quine: 'A Letter to Mr. Osterman'. Quine suggests that the boundaries between philosophy and science are arbitrary, like the boundaries between different districts on a map. But if this were true, agreement about new achievements as belonging to philosophy or to science would need to be conventionally settled; however, this is not the case. Such agreements seem to us natural and immediate.

[6] Ludwig Wittgenstein defended the view that philosophy is not constituted by argumentative theoretical conjectures, but is a therapeutic activity of describing how language really works (L. Wittgenstein: *Philosophische Untersuchungen,* sec. 109). Nevertheless, as many have remarked, neither Wittgenstein nor his followers have come remotely close to achieving this aim; the obscurity and elusiveness of Wittgenstein's arguments don't turn them into descriptions (see, for example, A.J. Ayer: *Wittgenstein*, p. 137).

of love (*philía*) and hate (*neíkos*) on the four elements (water, air, earth and fire), as the pre-Socratic philosopher Empedocles proposed, was an inevitably vague and obscure speculative dream. It was developed in a way and in a domain of inquiry in which researchers were not able to find a path to consensual agreement.

The conjectural character of philosophical thought – as the result of a lack of consensual agreement on fundamental matters – reveals itself as a necessary property, which is able to explain its typically *argumentative* and *aporetic* character. For when thinking cannot be other than conjectural, there is no alternative except to embark on hypothetical reasoning: We begin by accepting certain non-consensual assumptions and then apply our best knowledge and skills to discover all the implications of this acceptance. Then we (usually other philosophers do this) vary the assumptions and proceed in a similar way. And then we try to compare critically the different results and the procedures that lead to them, in a process that can be repeated again and again. In this way, philosophers are always beginning: they are always pondering new ideas in ways that usually generate aporetic argumentative discussion.

Moreover, in the conjectural character of philosophical inquiry we also find an explanation for the lack of progress in philosophy: since philosophers cannot achieve agreement on the truth of their ideas, inter-theoretical comparisons must remain inconclusive. (To give some examples: scientists would generally agree that Einstein's relativistic mechanics is superior to Newtonian mechanics, since the explanatory power of the former is greater – this is a matter of scientific conclusions. On the other hand, philosophers remain divided when the question is, e.g. whether the nominalism of the British empiricists might give a more suitable explanation for the problem of generality and predication than does some sort of Platonic-Aristotelian realism, for this is a matter of philosophical doctrine.)

Still, why can't we achieve consensual agreement about the results of the philosophical work? The answer is that *consensual agreement about the results of an investigation is only possible when there is sufficient agreement about the main presuppositions underlying the investigation.* A previous agreement about these things is always absent from philosophical inquiry. Philosophy lacks:

(i) agreement about the *adequacy of its data, principles and the questions based on them* (philosophical 'data' and principles are uncertain and many philosophical questions, we suspect, are pseudo-problems resulting from linguistic-conceptual confusions);

and philosophy also lacks

(ii) agreement about the *adequacy of its methodological procedures for evaluating the truth of answers proposed to philosophical questions* (an argument or a set of arguments can appear conclusive to one philosopher and irrelevant to another).

In opposition to this, conditions (i) and (ii) are always sufficiently satisfied in cases of scientific endeavours. For in science the questions and problems are uncontroversial, and the correct solutions, when finally found, can be clearly identified. Indeed, where fundamental conditions like these cannot be met, there is no way to achieve consensual agreement, and we remain doomed to the aporetic discussions typical of philosophy.

3 Philosophy as a protoscience

The foregoing remarks suggest that by investigating the similarities and contrasts between philosophy and science we may achieve a better explanation of some central features of philosophical inquiry. Moreover, it invites us to ask if our present philosophical inquiries will someday yield to science, by achieving a degree of maturity that allows its practitioners to reach consensual conclusions. In other words: could philosophy be seen as a conjectural inquiry anticipating science – as a *protoscience*? Could *all* philosophical inquiry be seen in this way?

A positive answer to this question is suggested by the historical fact that every new science was born in the cradle of philosophy. Consider some few examples from several scientific fields:

1. According to Karl Popper, the now obvious astronomical consideration that the earth is a body moving in empty space,

impelled by inertial and gravitational forces, was already anticipated by Anaximander, who suggested that the earth was a stationary cylinder, suspended in the void because it is equally distant from all things, it being impossible for it to move simultaneously in opposite directions[7].

2. The scientific investigation of subatomic particles by contemporary physics had as forerunner the speculative hypothesis of atomistic philosophers, from Democritus to Epicurus, that visible things are formed by the aggregation of invisible (because extremely small) and physically indivisible particles.
3. Biological theories of evolution seem to be dimly anticipated by Anaximander's insight that since man is helpless as a child, he would have perished in primeval times if he had not developed from more primitive animals…
4. The Platonic theory of the tripartite soul has a modern equivalent in Freud's structural theory of mind, which divides the mind into the *ego*, the *id* and the *superego*, although psychoanalysis still strongly resembles philosophy, insofar as its practitioners remain unable to reach consensual agreement.
5. Wittgenstein's therapeutic view of language as a nebula of language games working as unities of meaning anticipated the more scientific theory of speech acts of J.L. Austin and J.R. Searle, which nowadays belongs more to linguistics than to philosophy.

These are only a few examples. The process is going on in the present days. Many believe, for example, that as soon as we really understand how the brain works, most of the riddles of our present philosophy of mind will yield to consensual (and in this sense scientific) solutions. All these facts lead us to ask whether science might not someday replace the remaining central philosophical fields, such as epistemology, metaphysics and ethics.

Nevertheless, there are philosophers who resist the view of philosophy as a protoscience. Echoing Wittgenstein, Anthony Kenny holds that philosophy, unlike science, deals with knowledge as a

[7] This remark is made by K. R. Popper in 'Back to the Pre-Socratics', in his *Conjectures and Refutations*, p. 138.

whole, since it aims to organize the already known, providing an *overview* of our knowledge. This kind of *comprehensiveness*, he argues, is lacking in the individual sciences; central areas like metaphysics, epistemology, the theory of meaning and ethics, at least, will remain forever philosophical[8].

Nonetheless, an overview and comprehensiveness can be achieved by scientific inquiries too. Therefore, I suspect that the main reason for this resistance lies less in the nature of things than in outdated views of the nature of science still uncritically accepted by many philosophers. For these views, which have their roots in the philosophy of natural science developed by the positivists (and also in the main reactions against them), are often *too restrictive* to assure our central philosophical interests a deserved place in future scientific inquiry. Consider, for example, Popper's view of science as an inquiry that aims to create theories able to resist falsification by decisive experiments.[9] This view is too restrictive, even in the 'hard' domains of natural science: it excludes the theory of biological evolution from the domain of science, because it is not decisively falsifiable. How would we conceive an experiment capable of falsifying a hypothesis about a process that occurred in the past? Since we have this problem, how could such a restrictive standard as falsifiability (which may arguably only be applicable in certain fields of physics) ever be applied to the central subject matters of philosophical inquiry, like those of epistemology, metaphysics and ethics, other than in a crassly reductive or even eliminative way?

Indeed, were views like that the only available attempt to identify philosophy with a protoscience, our questioning could stop here. The reason for this is that from the investigation of a well-established particular science we are led almost perforce to restrictive generalizations about the character of still unknown areas of science. What we are looking for is a concept of science so general and inclusive that any new science that should chance to emerge can satisfy it, since this would be precisely the concept of science that we could properly contrast with that of philosophy.

[8] Antony Kenny: *Aquinas on Mind*, p. 9.

[9] See K.R. Popper, *Conjectures and Refutations*, *ibid.* pp. 339-340.

4 The right contrasting view of science

To arrive at this more balanced view of science we must follow the contrasting direction initially proposed and begin with questions like: what does the scientific community as a whole understand under the word 'science'? How would scientists recognize any new theory or field of investigation as belonging to science? I find such a balanced view of science in the work of John Ziman, who regarded science in general as 'consensualizable public knowledge', that is, as any kind of knowledge susceptible to consensual agreement concerning its results.[10] According to this view, science is constituted by generalizations consensually accepted as true by the members of a community of ideas. As Ziman shows, this notion has the advantage of being commonsensical: it is in perfect accord with what informed laymen and scientists understand as science. When we talk about science, we are thinking of new knowledge that the community of specialists can surely and precisely evaluate. This view of science is also sufficiently general and flexible to include *everything* we usually accept or could accept as belonging to the sciences, both the empirical and the formal. Moreover, placing the concept of consensual agreement at the center, Ziman's view of science seems to provide the perfect contrast between philosophy and science, since, as we have seen, the latter is an inquiry identifiable by a *lack* of consensual agreement concerning its results. Accordingly, even if philosophy might be a comprehensive inquiry aiming to achieve an overview, as Kenny thought, it might also be protoscientific, insofar as the possibility of becoming a sort of consensualizable public knowledge is not excluded in advance.

However, isn't the definition of science as 'consensualizable public knowledge' *too inclusive*? It seems to be so because there are political, religious and other communities in which a 'consensus' is imposed from above, excluding the possibility of critical evaluation. A notorious example of this is the role played by political ideologies in defining legitimate science in Nazi Germany and the Soviet Union. Yet, following the above characterization, such ideological impositions do

[10] This is the general thesis on the nature of science defended by J. M. Ziman in *Public Knowledge: An Essay Concerning the Social Dimension of Science*, chapter 2. For current serious discussion of the social dimensions of science, see J. M. Ziman, *Real Science*; see also H. L. Longino: *The Fate of Knowledge*.

seem to pertain to science, since a community of ideas consensually accept them. Thus, Ziman's characterization of science seems unable to distinguish science from a by-product of ideology.

Nonetheless, we see that this difficulty is only apparent, when we distinguish between *authentic* and *inauthentic* consensus, specifying what we understand as a community of ideas able to produce science in a way that excludes inauthentic consensus. Keeping the contrast with philosophy in mind, I suggest we call a community able to warrant authentic consensus a *critical community of ideas*, understanding it as a community which satisfies *constitutive conditions* approximating those specified by Jürgen Habermas for what he calls an *ideal speech situation* (*ideale Sprachsituation*)[11]. This means that we must define a critical community of ideas as something that satisfies constitutive conditions warranting authentic consensus. Without trying to be either systematic or exhaustive, I propose that we can generally characterize the main constitutive conditions for a critical community of ideas as requiring:

(a) *Truth-seeking commitment*: the members of the community should try to find the truth along the whole process of inquiry and evaluation of ideas.
(b) *Freedom of discussion*: there must be an equitable potential for free critical discussion among members of the critical community of ideas; they should not be subject to any intellectual constraints, except those of the best arguments.
(c) *Full access to information*: all members of the community must have full access to information and equal chances for the evaluation and exchange of ideas.
(d) *Shared competence*: all members must have suitable training in order to be able to make adequate evaluations.

Only the minimal satisfaction of constitutive conditions like these, assuring the freedom and the rationality of the evaluation of the results of the investigation, can produce a truly scientific consensus.

[11] See Jürgen Habermas: 'Wahrheitstheorien' in *Vorstudien und Ergänzungen zur Theorie des Kommunikativen Handelns*, p. 174 ff. See also H.L. Longino, *ibid.*, p. 128 ff.

In fact, when we evaluate reports of a new scientific discovery, we always do so on the assumption that the scientific community has satisfied the conditions of truth-seeking commitment, free discussion, full access to information, and shared competence, if not ideally, at least to a *sufficient* degree.

Another important objection that could be opposed to such a consensualist view of science is that it would compromise its objectivity. It seems at first glance that whatever the scientific community decides to call science will be science, arbitrarily disregarding objective criteria. However, this is not how things really are. For the critical community of ideas aims at a *consensus about truth*, and it cannot achieve this without sharing appropriated objectivity-conditions for the considered epistemic domain. In other words: it turns out to be an unavoidable fact of experience concerning the working of any community of ideas that consensus about truth is only attainable *after* the fulfilment of appropriate conditions for objective consensus. Therefore, without attempting to be neither systematic nor exhaustive, we can make a list of conditions that altogether must be satisfied by any object of evaluation by a critical community of ideas in order to warrant claiming the presence of the consensual objectivity necessary for the achievement of consensus about truth. This list requires that for the achievement of consensus about truth there must be at least previous agreement about:

(f) what can be counted as the (empirical or formal) *elementary data*, constituting the epistemic domains to which scientific theories are applied;
(g) what can be accepted as *adequately formulated questions* to be asked concerning the epistemic domain (theories must answer meaningful questions);
(h) what can be accepted as an *adequately constructed theory* relative to the epistemic domain (in its internal as well as in its external *coherence* within a wider conceptual framework); and
(i) what are accepted as the *procedures of truth-evaluation* relative to the epistemic domain of a theory (which involves some kind of *correspondence* between a theory and the facts the theory should explain, some kind of *verification procedure* for the achievement of this correspondence, etc).

The satisfaction of conditions like these is necessary for the objectivity of the scientific endeavour, coinciding in many ways with the kinds of things that philosophers of science often investigate in a detailed way. The difference is that although these philosophers have often regarded only such conditions, ignoring the social role of the critical community of ideas, we consider these conditions subsidiary to the appropriated functioning of a critical community of ideas. It is not an *a priori* matter, but an unavoidable matter of fact, seen by any critical community of ideas, that the members of will only achieve consensus about what is true insofar as these conditions are satisfied.

With the aid of these notions, we can improve Ziman's general characterization of science as 'consensualizable public knowledge'. Here is my proposal:

> THE SCIENCE = a body of non-trivial generalizations obtained by the members of a critical community of ideas, these generalizations being consensually held to be true by this community.

This seems to be a better view of science, regardless of what some philosophers might think. It is better because it fits well enough with what all people, from scientists to educated laypersons, understand by science in general. To apply the concept of science to any inquiry, our first requirement is the achievement of agreement about the truth of non-trivial generalizations among the members of a critical community of scientists. Only with this in hand can scientific progress take place.

5 The right contrasting view of philosophy

The above outlined consensualist-but-objectivist view of science allows us to establish an adequate contrast with philosophy, for now we can characterize the conjectural inquiry of philosophy by similarity and contrast with scientific investigation, as follows:

> THE PHILOSOPHY = a conjectural body of investigations aiming at the achievement of true generalizations by the members of a critical community of ideas (the community of philosophers), without this community being able to achieve any consensual

agreement on the truth of these generalizations.

Following this characterization, any conjectural inquiry in any domain of thought in which it is impossible to achieve a consensual body of truths can be regarded as of a philosophical nature. Its philosophical nature derives from the lack of satisfaction of the conditions (f)-(i) warranting consensual objectivity in a critical community of ideas. Indeed, in those difficult domains where science, understood as 'consensualizable public knowledge', remains impossible, only the conjectural inquiry of philosophy remains available. In this way, we can explain why philosophy, in conformity to the etymology of the word, is the *love* (*phylo*) of knowledge and wisdom (*sophia*) and not its attainment. In the words of Bertrand Russell: 'Science is what we know; philosophy is what we don't know'… 'Science is what we can prove to be true; philosophy is what we can't prove to be false'.[12] Indeed, when philosophy achieves consensual truth, it ceases to be philosophy and becomes science. Even the metaphilosophical view I am sketching in this paper can yield to science if the critical community of ideas achieves agreement on its truth.

Another point we should note is that the practice of philosophy always presupposes a critical community of ideas, even if in some cases (like those of Vico, Peirce and Nietzsche) in a contrafactual manner. A well-known charge against medieval philosophy is that by accepting Christian dogmatism as above criticism it fell short of satisfying this condition.

Finally, it could be objected that as a typically 'higher-order' form of inquiry, philosophical inquiry is by its own nature incapable of objective verification and, consequently, of the kind of objectively grounded consensus achieved by science. My response is that this view may well be overly pessimistic. The main reason to think so is that support for a theory is not only directly empirical – through verification – but also *inter-theoretical*. This kind of support can also be found in the sciences. Take, for example, the Darwinian theory of evolution. Darwin and his contemporaries developed this theory without recourse to genetics, since Mendel's work was unknown to early evolutionary theorists. Nevertheless, the subsequent rediscov-

[12] Quotes taken from Allan Wood's postscript to Bertrand Russell's *My Philosophical Development*.

ery of genetic theory by the scientific community provided extremely important inter-theoretical support for evolutionary theory. Something similar can also occur within 'higher-order' philosophical inquiry. In an exaggerated way it has already been suggested that the problems of philosophy are so intertwined that any problem will be only solved when all others have already been solved (Wittgenstein). Far from being pessimistic, this remark points to inter-theoretic support. Insofar as related fields of knowledge approach science, new inter-theoretical support for philosophical insights will be provided, paving the way for consensual scientific knowledge.

6 Protoscientific versus analytic-conceptual view

Once we accept the suggested views, we see that the supposedly essential differences in subject matter or even in method between philosophy and science are illusory. Take, for example, the still widely accepted view of the nature of philosophy which holds that it is a non-empirical, higher-order activity of *conceptual analysis* (its method), intended *to make explicit the structure of our most central concepts and the relations holding between them* (its subject matter)[13]. This view arose due to the prominence of the philosophy of language in the first half of the Twentieth Century. But it was factually refuted when the philosophy of language, as the most productive philosophical field, was superseded by the philosophy of mind in the second half of the Twentieth Century, since the latter philosophical field consists largely of *empirical* speculation. Moreover, the fact that a given philosophical inquiry has a linguistic-conceptual character does not mean it cannot develop into a science. This is exemplified by J. L. Austin's theory of illocutionary forces. As he himself foresaw, this theory belongs today, in the form of the theory of speech acts, more to the scientific field of pragmatics than to philosophy. And the reason for this is that it has achieved enough consensual agreement to lose its plastic role

[13] The persistence of this view can be exemplified in the essays of Robert Brandom, Barry Stroud, Allen Wood and Karl-Otto Apel, recently published by C.P. Ragland & S. Reidt in *What is Philosophy?* For standard presentations of the view see Michael Dummett: 'Can Analytical Philosophy be Systematic and Ought it to Be?' and Ernst Tugendhat 'Überlegungen über die Methode der Philosophie aus Analytischer Sicht'.

in the domain of conjectural thought. Hence, there seems to be no contradiction between the view of philosophy as a protoscience and the view of philosophy as conceptual analysis, since the latter can be regarded as belonging to the former[14].

Finally, we can offer a metaphilosophical refutation of the thesis that the *proper object* of philosophy is conceptual. As W. v. O. Quine saw, philosophers often need the resource of what he called a *semantic accent*.[15] A semantic accent is what we could also call a *semantic metalanguage*. A semantic metalanguage is something different from a syntactic metalanguage: while the latter has as its objects signs and their relationships, the former has as its objects as well the meanings and with them, indirectly, *the world as we mean it*. (For example, instead of saying, 'Five is not a thing but a number', one would say, ''Five' is not a thing-word, but a number-word'). This need not lead us astray, for it is nothing more than a *propaedeutic* resource aiming at the achievement of the kind of conceptual transparency usually demanded by philosophical arguments. Even when philosophers like Rudolph Carnap have seen here a proof that the object of philosophy should be purely linguistic-conceptual, this cannot be true, as Quine also noted, because every sentence of the empirical sciences can also be metalinguistically represented in this way. As he noted:

> There are wombats in Tasmania' might be paraphrased as ''Wombats' is true to some creatures of Tasmania' if there were any point in it; but it does happen that semantic accent is more useful in philosophical connections.[16]

The upshot of this is that philosophy does not have concepts (like those of meaning or knowledge or consciousness or substance) as its proper subject matter, any more than science (with concepts like those of genes, molecules and superstrings), and no more than any-

[14] J.L. Austin also saw no contradiction between philosophy as protoscience and philosophical analysis, since on the one hand he championed philosophy as conceptual analysis and on the other hand he was an inspired defender of the here-developed view (see the much quoted passage of 'A Plea for Excuses', in his *Philosophical Papers*, p. 232).

[15] W.v.O. Quine, *Word and Object,* p. 270 ss. See also my *The Philosophical Inquiry*, pp. 15 ff.

[16] W.v.O. Quine: *Word and Object*, pp. 271-272.

thing else, except for reasons of semantic accent. A theoretical physicist and a philosopher of mind, for example, can be seen as analyzing and combining concepts as much as dealing with empirical questions. Hence, all that we can intend by saying that philosophy is conceptual analysis is to point to some methodological resources, and not to an indispensable approach, and still less to its proper subject matter.

To the question of whether all philosophy might be an anticipation of science, assuming the concept of science that we have suggested, the only answer is that we have no reason to think otherwise. In fact, the only chance we have to make real progress is by holding this view as a normative assumption.[17]

7 The more complete framework

While I have limited myself here, for methodological reasons, to the relationship between philosophy and science, I believe that this is only one aspect of the more complete framework that places philosophy within a broader perspective. In my book on the nature of philosophy.[18] I attempted to achieve this broader perspective conceiving traditional philosophy as a derivative cultural activity that can be seen as an amalgam of three other cultural activities: art, religion, and science. These activities could be represented as forming the edges of a triangle inside of which different philosophical activities can be placed. The scientific edge of the triangle is success and truth-oriented, the mystical-religious edge is responsible for the transcendental element, and the characteristic breadth of the philosophical quest, the aesthetic-artistic edge, is responsible for the metaphorical aspects of the philosophical discourse. Inside the triangle, in its centre, can be found philosophies that have in

[17] There are many further problems that cannot be touched on here. For example, how to include certain non-central domains, like those of the philosophy of existence, philosophy of life, or critique of culture – which have changeable subjects – in our schema? (Probably in ways similar to those in which the historical sciences can be consensualizable). Another point is that the development of science can itself create a space for new and previously unexpected philosophical fields. The philosophy of computational science is an example.

[18] C.F. Costa: *The Philosophical Inquiry: Towards a Global Account*.

sufficient measure artistic (metaphorical), religious (transcendental) and scientific (truth-oriented) aspects (examples are Platon's *Republic*, Descartes' *Meditations*, Wittgenstein's *Tractatus*). However, there are forms of philosophy that are located near the artistic edge (like Nietzsche's *Zaratustra* and Novalis' *Hymns to the Night*), near to the mystical-religious edge (like John Scott's *The Divisions of Nature* and Meister Eckhart's *Sermones*, near to the scientific edge (like Searle's *Speech Acts* and Carnap's *Logical Grammar of Language*). There are even cultural traditions linking philosophy with an edge, like the French literature oriented tradition, the German mystically oriented tradition and the English scientifically oriented tradition. And it is possible to perceive in the sub-domains a broad slow movement from the artistic-mystical side of the triangle to the scientific edge as an inevitable consequence of the speeding progress of science.

Cláudio F. Costa
Federal University of Rio Grande do Norte

References

Austin, J.L. A Plea for Excuses. In *Philosophical Papers.* Oxford: Oxford University Press, 1961.
Ayer, A.J. *Wittgenstein.* New York: Pelican Books, 1985.
Brandom, R., Barry Stroud, Allen Wood and Karl-Otto Apel. In *What is Philosophy?* ed. by C.P. Ragland & S. Reidt. Yale: Yale University Press, 2001.
Comte, A. *Cours de Philosophie Positive.* In *Ouvres.* Paris: 1968 (1830-1842), vol. 1.
Costa, C.F. 2002. *The Philosophical Inquiry: Towards a Global Account*. Langham: UPA.
Dummett, M. Can Analytic Philosophy be Systematic and Ought to Be? In *Truth and Other Enigmas*. London: Duckworth, 1978.
Guthrie, W.K.C. *A History of Greek Philosophy*. Vol. 1. Cambridge: Cambridge University Press, 1962.
Habermas, J. Wahrheitstheorien. In *Vorstudien und Ergänzungen zur Theorie des Kommunikativen Handelns*. Frankfurt: Suhrkamp, 1984.
Kenny, A. *Aquinas on Mind*. London: Routledge, 1993.
Longino, H.L. *The Fate of Knowledge*. Princeton: Princeton University Press, 2002.
Odell (eds.) *The Owl of Minerva*. New York: Harvester Press, 1975.
Popper, K.R. Back to the Pre-Socratics. In *Conjectures and Refutations*. London: Routledge, 1962.
Quine, W.v.O. A Letter to Mr. Osterman. In *The Owl of Minerva*, ed. by C.J. Bontempo & S.J. Odell. New York: Free Press, 1975.
Quine, W.v.O. *Word and Object*. Cambridge: MIT-Press, 1960.
Russell, B. *My Philosophical Development*. London: Routledge, 2004.
Strawson, P.F. Analysis, Science and Metaphysics. In *Linguistic Turn: Recent Essays in Philosophical Method*, ed. by R. Rorthy. Chicago: Chicago University Press,

1967.
Tugendhat, E. Überlegungen über die Methode der Philosophie aus Analytischer Sicht. In *Philosophische Aufsätze*. Frankfurt: Suhrkamp, 1992.
Wittgenstein, L. *Philosophische Untersuchungen*. Frankfurt: Suhrkamp, 1986.
Wood, A. postscript of B. Russell's *My Philosophical Development*. London: George Allen & Unwin, 1975.
Ziman, J.M. *Public Knowledge: an Essay Concerning the Social Dimension of Science*. New York: Cambridge University Press 1968).
Ziman, J.M. *Real Science*. Cambridge: Cambridge University Press, 2000.

On the Notion of Object. A Logical Genealogy

Fernando Ferreira
University of Lisbon

BIBLID [0873-626X (2012) 34; pp. 609-624]

1 Introduction

The most general notion of object has its modern home in the first-order classical logic with identity. In this paper, we argue that the apparatus for speaking of objects finds its proper place among a wider logical setting. The purpose of this paper is to explain and defend this thesis. One of the benefits of our analysis is that it makes possible to isolate the logical principles that are characteristic of the notion of object.

The wider apparatus of logic is brought about in virtue of two contentions. The first one is that hypothetical and general statements are the fundamental and primitive notions that make reasoning possible. In section 5, we briefly try to argue for this claim. Our argument relies on a fastened view between logic and inference. The second contention is that concept quantification is a coherent and admissible form of logical expression. Concept quantification has been famously attacked, either for being incoherent, or as a form of (extensional) second-order quantification lying outside the province of logic. After a brief discussion of conventionalism in logic in section 2, we discuss some theses of Willard Quine that are pertinent to our discussion. This is done in section 3 and, forthwith, a defense of a form of second-order logic is mounted. In the following section 4, we make a small digression on the principle of induction and on the benefits of its enunciation with concept quantification.

In sections 6 and 7, we finally describe the logical principles that articulate the notion of object. A small concluding section closes the paper.

2 Conventionalism in logic

There is an interesting, albeit failed, account of logic. It is the view that explicates logic as the adoption of certain linguistic conventions. For instance, adopting (or not) *tertium non datur* would be a matter of convention within a particular linguistic framework. The conventionalist strategy can be compared to the discovery of non-Euclidean geometries. Denying the parallel postulate is not a contradiction in terms or a failure to describe correctly some reality, but rather adopting another kind of geometry. If mathematicians came to accept different geometries side by side, why shouldn't philosophers accept also different logics side by side? Why can't we be free to adopt different logics, different ways of reasoning, in order to pursue more efficiently some inquiries?

A conventionalist view of logic would, in one single sweep, explain why the truths of logic are of a different kind from empirical truths, and why they are analytic and *a priori*. Of course, this would be an important step in the enterprise of logical positivism. A logic would merely be an adopted calculus where, by conventional stipulation, some inferences are permitted and some principles are asserted. The stipulations would be seen as implicit definitions of the terms of the calculus.

Conventionalism in logic was refuted by Quine more than seventy years ago. The difficulty lies in that logic itself is presupposed in establishing non-trivial conventions.[1] Suppose that we want to establish conventions for the *if*-idiom. We may start by laying down the following convention:

> *If x and z are true sentences, y is a sentence and z is the result of substituting the letter 'p' for x and the letter 'q' for y in 'If p then q', then y is true.*

Note, however, that the convention uses the *if*-idiom itself, as well as the *and*-idiom and, implicitly, the *all*-idiom. If we do not understand these idioms beforehand and know how to operate with them, the above convention does not get off the ground. I include this Quinean analysis here because, later in the paper, I will argue that both the *if*-

[1] See specially part III of Quine 76. Cf. also the short and witty article by Carroll 1895.

idiom (the idiom of hypothetical claims) and the *all*-idiom (the idiom of general claims) are the fundamental and primitive notions which make logic possible.

3 Concept quantification

The above argument of Quine was one of his first attacks on logical positivism, an attack which culminated in the famous *Two Dogmas of Empiricism* paper (see Quine 1953b) and its attendant rejection of the analytic-synthetic distinction. However, and this is important to notice, the paper is actually only an attack on a certain class of analytic statements, viz. the ones that can be turned into a logical truth by putting synonyms for synonyms. The class of logical truths is not itself subjected to attack in Quine's paper. Quine's real target is the reification of the notion of meaning, an attack mounted through the notion of synonymy. Presumably, if the notion of synonymy were acceptable, meanings could be reified: indeed, they could be defined as equivalent classes of terms under the relation of synonymy. The attack on synonymy had great success but, as we will see, combined with another thesis of Quine, had an unfortunate and important collateral damage.

There are many problems with the notion of reified proposition (as meanings of sentences) but Quine is famous for rejecting an ontology of propositions on the basis that it is hard to account for their identity conditions. It is part and parcel of the notion of object that the *same* object can be referred to by means of different descriptions or from different perspectives. Identity, as stressed by Gottlob Frege, is an important part of the apparatus of linguistic reference to objects. Quine also stresses quantification. At any rate, the most general notion of object finds its modern home in first-order classical calculus with identity. If one takes seriously the view that the notion of object is a quintessential metaphysical notion, then first-order classical logic with identity must be deeply ingrained with metaphysical presuppositions. I will recount a 'logico/metaphysical story' on how the notion of object is brought into logic. The aim is to make explicit, via inferential articulations, the metaphysical presuppositions of this notion.

Speaking of objects, as it is remarked in Parsons (2008: 10), just

is using the linguistic devices of singular terms, predication, identity and quantification to make serious statements. Meanings and propositions do not qualify as objects. According to Quine, they do not so qualify because they miss at least one of the requirements, the one on identity. However, as Quine 1953a is careful to point out, he does not deny that words and statements are meaningful. He just denies that they *have* a meaning, if this 'having' is interpreted as more than a mere figure of speech. We, humans, *use* meaningful statements all the time. The point is that using them is not the same thing as naming them, nor does it presuppose that we can name them (i.e., that propositions can be treated as objects).

However, Quine goes a step too far when he defends that quantification is not only a *necessary* condition for speaking of objects but that it is also a *sufficient* condition. *To be is to be the value of a bound variable*, in Quine's famous, but ill-conceived, dictum. Quine argues for this thesis in *On what there is?* (Quine 1953a: 15) where he also lays down his doctrine of ontological commitment. The argument of Quine proceeds via an analysis of complex descriptive names which do not refer. He analyzes these names via Russell's theory of descriptions and points out that the ontological commitment is carried out by the bound variables. Notice, however, that Russell's theory relies on the full apparatus of first-order classical logic with identity: it needs predication, quantification and identity. Therefore, what is argued by Quine is that in the presence of this linguistic apparatus (the apparatus for speaking of objects), ontological commitment is carried out by the bound variables. So far, so good. However, to conclude from this argument that bound variables carry ontological commitments in the absence of the full apparatus is a *non sequitur*.

Quine's emphasis on quantification is right in one direction - speaking of objects presupposes the availability of the apparatus of quantification (with ontological commitment) - but wrong-headed in the other - that the mere use of the apparatus of quantification signals ontological commitments. The collateral damage is, of course, that quantifying over propositions (or, in general, over concepts) does not make sense. As I said above, I will present a 'logico-metaphysical' story on how the notion of object is brought into logic. In this story, the notion of generality is primitive and conceptually prior to the notion of object and, by itself, does not signal ontological commit-

ments. It is merely a means of expression.

Before I tell my story, let me make a few remarks. I start with an example: '$\forall P(P \rightarrow P)$', where 'P' is a second-order propositional bound variable. This expression, I submit, makes perfectly good sense.[2] On our view of propositional generality, an instantiation of a universal (second-order) propositional quantification is *not* obtained by converting each occurrence of the quantified variable into an expression that *names*, but rather into an expression that *propounds* (i.e., a meaningful sentence). That which can be propounded is essentially open-ended and unfinishable, depending on linguistic and conceptual resources of whose limits we have no real conception. It is impossible to survey all propositions or concepts because 'the attempt to survey reason itself fails: reason can transcend whatever it can survey' (Putnam 1998: 119).

It is important to point out that the role given to meaningful sentences in the conclusion of the elimination rule for universal propositional quantification is one in which they appear in positions of *use*. Let me give an example: 'Bustopher is a fat cat $\rightarrow$ Bustopher is a fat cat' is an instantiation of '$\forall P(P \rightarrow P)$'. Do notice that the expression 'Bustopher is a fat cat' appears in positions of use in the sentence 'Bustopher is a fat cat $\rightarrow$ Bustopher is a fat cat'. Even though the notion of proposition does not have the right content to allow proposi-

[2] During a presentation of this work, I mentioned that it makes perfectly good sense to say that every proposition implies itself. With his customary politeness, Ricardo Santos pointed to me that I probably did not want to use the word 'implies' here. Implication is reserved for inference, as when one says that the sentence 'The cat is on the mat and the dog is in the garden' implies the sentence 'The dog is in the garden'. The word 'implies' is part of the metalanguage and cannot function as a sentential operator. For this function, one should use instead the conditional 'if … then …' as in 'If the cat is on the mat and the dog is in the garden then the dog is in the garden'. Of course, this gives rise to a *grammatical* problem: to say 'every so is such that if so then so' is not proper English. Locutions like 'every' or 'for all' demand to be followed by noun phrases. Grammar pushes us to say something like 'every proposition is such that if it holds, then it holds'. In a nutshell, in our example, grammar demands ordinary language to treat propositions as objects (and truth as a property thereof). This grammatical objection is taken very seriously by some authors. See, for instance, the critique of Burgess (2005: 211ff). Grammar and 'common usage' do not have, however, a good repute as a guide to philosophical or scientific inquiry (see the caustic Russell 1953).

tions to function *as objects* in the range of a quantification, I am arguing that it has a content which permits their use within the *all*-idiom in the manner which I have just described. In actual instantiations, the question of whether a sequence of symbols counts as expressing a proposition can only be answered by way of interpretation and public agreement. It is, irredeemably, a matter of interpretation. It is certainly right to say that, in meaningful exchange, interpretation always lurks.[3]

4 Digression on the natural numbers

Given the coherence of propositional and concept quantification, a refusal to accept this kind of quantification is tantamount to a prohibition to engage in concept building and expression. It is an unwillingness to go on in the direction of a greater linguistic expressiveness. Let me make a brief digression and discuss an example from mathematics: the principle of mathematical induction:

$$\forall H\,[H(0) \wedge \forall n(H(n) \rightarrow H(n+1)) \rightarrow \forall n H(n)]$$

The use of this principle is, directly or indirectly, behind almost all of the mathematics that is indispensable for science. In the light of the discussion of the last section, I purposely enunciate the principle with a (second-order) concept quantification because this is a way to express its universality, in the sense that the properties to which the principle of induction applies are not parochial to a fixed language. The principle rather applies to properties whose expression is part of

[3] The notion of propositional quantification has some superficial affinities with substitutional quantification. There are two crucial differences, however. The instantiations of a universal propositional quantification are not obtained from sentences of a given fixed formal language. As a consequence, propositional quantification is not amenable to semantic ascent and cannot be reformulated as quantification over a certain fixed domain of sentences together with the concurrence of a truth (or satisfaction) predicate. Moreover, in our notion of quantification, sentences formed using the very apparatus of propositional quantification can themselves be used as instantiations of universal quantifications. In other words, contrary to substitutional quantification, propositional quantification accepts impredicativity and, when generalized to concept quantification, it allows for the formation of impredicative concepts.

an open-ended process.

The expert reader knows that the mathematics indispensable for science only requires enunciations of the principle of induction within particular formal languages.[4] We can certainly opt for this kind of parochialism with respect to science as it is currently practiced. Nevertheless, the principle as enunciated above not only has a regulative force - in the sense that the evidence for the truth of its formal (schematic) enunciations are best seen as coming from the general principle[5] - but it also sheds light on the relationships between provability and truth in mathematics.

Arithmetic, when developed in a formal language, is subjected to Gödel's first incompleteness theorem. According to this theorem, under mild technical conditions, a consistent theory of arithmetic is necessarily incomplete, i.e., there are always sentences of the formal language which are neither provable nor refutable (their negations are not provable). If we further see the system of arithmetic as aiming at proving true arithmetical sentences, we can say that there are always true sentences of arithmetic which are not provable in the given formal system.

The proof of Gödel's theorem relies essentially on the complete formalization of the arithmetical theory. If the principle of induction is formulated as above, with concept quantification, Gödel's theorem simply does not apply. This state of affairs should not come as a surprise because in any formal deductive system the statement of the principle of induction is, by necessity, *restricted*.[6] Not only is this

---

[4] For instance, in the language of set theory. Mathematical logicians have studied in detail what exactly is necessary - in terms of induction or set comprehension - for proving ordinary theorems of mathematics (for a reference, see Simpson 1999). For the connection of this kind of work with science and the indispensability arguments, I recommend Feferman 1998.

[5] This simple point was given its due weight by Kreisel (1967: 148).

[6] The observation about the open-endedness of concept formation is quite plain. Ditto for the observation that the principle of induction is open-ended. However, the following question and (especially) its answer are rather deep. Let be given a formal language of arithmetic. It has a delimited apparatus of concept formation and, therefore, a restricted amount of induction. Would it nevertheless be possible to set up a (consistent) theory in the formal language that would either prove or refute each sentence of that very same formal (and, hence, restricted)

observation in line with the open-endness of concept formation, but it also finds rigorous mathematical support in the theorem that the truth predicate for a formal language of arithmetic cannot be defined within the language itself.[7]

Wherein is exactly located the presumed gap between truth and provability intimated by Gödel's first incompleteness theorem? I suggest that the gap, if it is a gap, lies in the open-endness of concept formation and, in our case, in the open-endness of numerical concept formation.

5 The centrality of hypothetical and general statements

Inference is central to logic and reason. Without being able to move from premises to a conclusion, there is no reasoning. Logical truth is a particular case (or a degenerate case, as when we say that a point is a degenerate circle) of inference: one without premises. The degenerate case is no substitute for the general case since logical truth alone does not account for the *moves* from statements to statements which characterize logic. It is an obvious point that it is impossible to set up a logical system without at least one inference rule. Without inference, there are only isolated proclamations.[8]

Suppose that the primitive man concludes that he is in danger from the information that there is a lion in the vicinity. A reflection on what he *does*, i.e., the *move* of concluding that he is danger from the knowledge that there is a lion in the vicinity, and the will (and capability) to express this movement, takes him to say 'if there is a lion in the vicinity then I am in danger'. Hypothetical (or conditional) statements express, within language, the *sanction* of the linguistic act of drawing a conclusion under certain conditions. They are a fundamental form of logical expression and make possible a linguistic ascent, from linguistic act into linguistic expression, without which

---

language? Gödel's first incompleteness theorem says that this is impossible.

[7] This is Tarski's undefinability of truth theorem.

[8] Such proclamatory discourse would marry well with the notion of language as consisting only of *descriptions of reality*. Interestingly, in Sellars 1997 it is argued that even perceptual statements can only count as assertions insofar as they are inferentially articulated.

there is no explicit reasoning.[9,10]

In the example above, a measure of *generality* is implicit. The conditional sentence 'if there is a lion in the vicinity then I am in danger' operates like an *inference ticket* which can be cashed in *all* the appropriate situations.[11] In the complete absence of generality, even though there is more in discourse than mere isolated proclamations, we are - in a sense - only a finite number of steps away from inferring all that can be concluded. Let us advance an argument for this thesis. Arguably, the main role of the elimination rules of a connective is to make possible the use of premises in which this connective occurs as the principal connective. It is plain that if the conditional is the *only* logical connective present in the premises of an argument, we would be only a finite number of steps away from inferring all that could be inferred via the elimination rule (i.e., via *modus ponens*).[12,13] On the

[9] The content of hypothetical statements results from some distinguished roles in inference. More precisely, these are the roles carried out by the introduction and elimination rules of the natural deduction calculus. This calculus was invented by Gerhard Gentzen in 1935, and subsequently studied and expanded in Prawitz 1965.

[10] The attentive reader can point that there are inferences with more than one premise, and that this fact calls for the notion of conjunction as well. This is correct, but not terribly interesting. There is a further factor that explains our simplified account: technically, as it will be pointed out, conjunctive claims can be expressed using hypothetical and general claims.

[11] This point was made a long time ago by Ryle 1950.

[12] The above argument is not as strong as the argument for the centrality of hypothetical statements. The caveat 'in a sense' is needed because the introdution rule for implication does allow the inference of infinitely many sentences (uninteresting as they may be). Nevertheless, we believe that it carries a certain weight. It would be nice to advance stronger and more perspicuous arguments for the centrality of general statements.

[13] Frege 1984 also assigns a central role in logic to generality. Frege speaks of the scientific need to express *laws* and says that 'in point of fact the distinction between law and particular fact cuts very deep'. Frege, on the other hand, does not accord a centrality to hypothetical statements. He sees them as truth functional and, therefore, replaceable by the combination of negation and conjunction (cf. Frege 1980). Of course, Frege is right: truth-functional conditionals (as any truth functional connective) can even be written with the Sheffer stroke only. However, our analysis of truth-functionality is not yet in place. Our rendition of

other hand, generality releases reasoning from a pre-fixed enclosure into a boundlessness of inferential moves.

The sentence '$\forall P(P \to P)$' is an example of a (true) general sentence. Notice that the number of immediate inferences which can be drawn from it (via the elimination rule) is unbounded. General claims, as hypothetical claims, are also a primitive and unexplained notion, fundamental to logical expression. These claims can be made not only with respect to predicate positions, but also with respect to name positions. However, it is methodologically interesting to start with propositional generalities - as we did - because doing so only requires the notion of proposition (i.e., of a meaningful sentence) and does not rely yet on any particular analysis of this notion.

At this juncture, it is worth remarking that the familiar propositional connectives can be introduced with the apparatus of hypothetical and general propositional claims:[14]

$$\begin{aligned} \neg A &=_{df} A \to \forall P.P \\ A \wedge B &=_{df} \forall P\,((A \to (B \to P)) \to P) \\ A \vee B &=_{df} \forall P\,((A \to P) \to ((B \to P) \to P)) \end{aligned}$$

These are the *expressive* definitions of the above connectives.[15] It is known that the attendant notion of consequence gives rise to intuitionistic logic.[16]

logic is still on a prior and more general footing.

[14] The definitions appear informally in sections 18 and 19 of Russell 1996. The formal definitions are due to Prawitz 1965.

[15] For the reader unfamiliar with these definitions, let us discuss the definition of disjunction. In natural deduction, the elimination rule of disjunction is the following: If 'P' can be inferred from 'A' and if 'P' can be inferred from 'B', then 'P' is a consequence of 'A∨B' (this is also known as *discussion by cases*). Given that concept quantification is allowed in our language, the definition of disjunction above mirrors the elimination rule just described. *Mutatis mutandis* for the other connectives. The introduction rules of the natural deduction calculus follow from the definitions.

[16] The notion of consequence is the one originating from the rules of introduction and elimination (for the conditional and the universal quantifier) of the natural deduction calculus (cf. footnote 9). It is a mathematical observation that these rules give rise to intuitionistic logic.

6 The *alethic* ingredient

The previous section recounts the first part of our 'logico/metaphysical story'. The subsequent story narrates the appearance of the notion of object and, in general, of the descriptive component of logic. We have already introduced the expressive ingredient of logic.[17] Let us now introduce two other ingredients: firstly, the *alethic* ingredient (i.e., the logical apparatus for the notion of bivalent truth) and, in the next section, the *ontic* ingredient (i.e., the logical apparatus for the notion of object).

The *alethic* ingredient of logic only comes into play when we are dealing with a linguistic base whose semantics is taken to be unproblematically bivalent. We start with a initial bit of language whose propositions are true or false, but not both (the principle of bivalence). For convenience, let us call them atomic propositions.[18] We take that atomic propositions come in pairs, one called the *opposite* of the other (with switched truth values). The opposite of an opposite is the original atomic proposition. This is our rendition of bivalence. The *inferential explication* of bivalence can be done through the following laws:

$(\mathsf{ALE})_{1\text{-}}$ $\forall P((A \rightarrow P) \rightarrow ((\underline{A} \rightarrow P) \rightarrow P))$
$(\mathsf{ALE})_{2\text{-}}$ $\forall P(A \rightarrow (\underline{A} \rightarrow P))$

where the letter 'A' stands for a sentence that expresses an atomic proposition (an atomic sentence), and '$\underline{A}$' stands for its opposite. The first law is a form of *tertium non datur* and can be elucidated by saying that an atomic proposition and its opposite exhaust the space of logical possibilities in the following sense: Whenever we want to infer a given proposition, it is enough to infer it from an atomic proposition

[17] Generalizations with respect to name positions (which will be discussed below) are also part of the apparatus of the expressive component of logic, but they presuppose a previous analysis of the structure of predicative statements. For the record, the expressive definition of '$\exists x\, A(x)$' is the formula '$\forall P(\forall x(A(x) \rightarrow P) \rightarrow P)$'.

[18] This is just a convenient way of speaking. There are no metaphysical connotations in the use of this terminology (not, for instance, connotations to Tractarian *Elementarsätze* or semantical atomism) beyond what we have said: atomic propositions are either true or false, but not both.

and to infer it from its opposite. The second principle is a form of *ex falso quodlibet*.[19]

If 'A' stands for an atomic sentence, it is easy to prove both that $\neg A \leftrightarrow \underline{A}$ and that $\neg \underline{A} \leftrightarrow A$.[20] In other words, for atomic propositions, expressive negation is equivalent to (bivalent) opposition. Furthermore, if we consider the sentences of the language built up from atomic sentences by means of propositional connectives (the conditional, and negation, conjunction, disjunction, as defined above) - originating *propositional sentences* - it is easy to argue by induction on the build up of sentences, that the statements of the form

$$\forall P\,((A \to P) \to ((\neg A \to P) \to P))$$

where 'A' now stands for a propositional sentence, are provable. Note that the above says that $A \vee \neg A$. It is now clear that the propositional fragment of the language obeys the laws of classical logic.

Our definitions of the propositional connectives are done in terms of propositional quantifications. It may cause some uneasiness the fact that the apparatus of propositional quantification is present in the propositional calculus (via the very definitions of the propositional connectives). The uneasiness is, however, uncalled for. In a sense, the apparatus of propositional quantification present in the midst of the propositional fragment is trumped by the alethic principles. The familiar (self-enclosed) set up of the propositional calculus via truth tables may, on this view, be considered an 'in your face' manner of displaying the *result* of this trumping.

In the same vein, it is worth remarking explicitly that the alethic trumping mentioned above explains the compositional nature of the classical propositional calculus. The like-minded reader must have noticed that the semantics of propositional quantification cannot be fully compositional. The reason for this lies in the impredicativity of concept quantification and the attendant consequence that one must

[19] With the definitions of the propositional connectives above, the sentence in $(ALE)_1$ is equivalent to '$A \vee \underline{A}$', and the sentence in $(ALE)_2$ is equivalent to '$\neg(A \wedge \underline{A})$'.

[20] The claim that $A \leftrightarrow B$ should be understood as an abbreviation for the pair of claims that $A \to B$ and that $B \to A$.

deny that the sense of a second-order generalization is intelligible only if its instantiations are intelligible in advance of the generalization itself.[21] This observation effectively precludes a compositional semantics *in general*. However, in the presence of the alethic principles, the restricted area of the propositional calculus is fully rooted in its bivalent beginnings, disarming - as it were - the role of the concept quantifications in its midst and, therefore, making it possible to cohere with a (compositional) truth-functional semantics.

7 The *ontic* ingredient

In the previous section, we brought into discussion a linguistic base whose semantics was taken to be unproblematically bivalent. For convenience, we called the elements of this linguistic base 'atomic propositions'. The expressive devices of hypothetical and generality claims create their own cargo by permitting the expression of new propositions which, themselves, can be subjected to the expressive devices again. And so on. A quite *diverse* cargo is created. Even if the atomic base has an unproblematic bivalent semantics, the application of these devices need not maintain language within such bivalent confines. What we have argued in the previous section is that they do if we restrict ourselves to the propositional calculus.

In this section, we further suppose that these atomic propositions have a predicative structure. For instance, we may say that Bustopher is a fat cat or that Mungojerrie and Rumpelteazer are twin cats. Proper names and predicate symbols mark places in sentences which are suitable for generalizations. They indeed mark different sorts of places, since proper names do not fit into predicate places nor vice versa. A sentence of the form '$\forall H(H(\beta) \rightarrow H(\beta))$, is a generalization of the sentence 'Bustopher is a fat cat → Bustopher is a fat cat', where 'β' denotes Bustopher. It is a generalization with respect to the predicate place (a concept quantification). We could further generalize in

[21] A second-order generalization may have instantiations which are not simpler than the generalization itself since the bound second-order variable may be instantiated by formulas of arbitrary complexity. In this essay we are not discussing the semantics of concept quantification but it is clear that it must rely on a species of a rule-following semantics. The reader can find an entertaining discussion of impredicativity in Ferreira 2006.

the name place and write '$\forall x \forall H(H(x) \rightarrow H(x))$'. First-order quantification is also a device for the expression of generalizations, through which a new manner of formation of concepts is allowed.

In order to deal with the first-order classical predicate calculus, we must extend our analysis of the atomic case to atomic *formulas*. In analogy with the last section, if the letter 'A' stands for a predicative symbol (unary: to simplify), then '$\underline{A}$' stands for its opposite. The extended alethic principles take the form:

$(ALE)_1$ $\forall x \forall P((A(x) \rightarrow P) \rightarrow ((\underline{A}(x) \rightarrow P) \rightarrow P))$;
$(ALE)_2$ $\forall x \forall P(A(x) \rightarrow (\underline{A}(x) \rightarrow P))$.

Disregarding the identity axioms, the *inferential explication* of the notion of object can completed through the following principle:

$$(ONT) \quad \forall M \forall K [\forall x \forall P((M(x) \rightarrow P) \rightarrow ((K(x) \rightarrow P) \rightarrow P)) \rightarrow \forall P((\forall x M(x) \rightarrow P) \rightarrow (\forall x (K(x) \rightarrow P) \rightarrow P))].$$

Here is an elucidation of this principle: If for each object x, reasoning from the fact that x falls under M and from the fact that x falls under K exhausts the space of logical possibilities, then reasoning from $\forall x M(x)$ and from a generic example falling under K also exhausts the space of logical possibilities.

Arguably, the principle (ONT) is a bit opaque on a first reading,[22] but we can frame it in a more familiar terminology: With the definitions of the propositional connectives and of the existential first-order quantifier in footnote 17, (ONT) is equivalent to

$$\forall M \forall K [\forall x (M(x) \vee K(x)) \rightarrow \forall x M(x) \vee \exists x K(x)].$$

It is also worth remarking that if there are only finitely many objects in the range of x, (ONT) is provable. For instance, if there are only two objects α and β, the principle takes the form

$$\forall M \forall K [(M(\alpha) \vee K(\alpha)) \wedge (M(\beta) \vee K(\beta)) \rightarrow (M(\alpha) \wedge M(\beta)) \vee K(\alpha) \vee K(\beta)].$$

[22] There is, however, no reason why the inferential explication of the notion of object must be prosaic.

The rendition of the notion of object is given by the inferential rules which regulate the use of the 'linguistic devices of singular terms, predication, identity and quantification'. (ONT) is part of this rendition. If I were pressed to elucidate its particular role in the notion of object, I would say that (ONT) conveys the notion of quantification over a *closed totality*.[23]

It is a simple exercise (via an inductive argument on the complexity of formulas) to show that (ONT), together with the alethic principles, entails that first-order formulas satisfy *tertium non datur*.[24] The principles of classical logic for the first-order fragment follow suit.

8 Coda

Discourse about objects finds its modern home in first-order classical logic with identity. The main thesis of this paper is that logic is not a uniform terrain where all truths lie on a par. We have analyzed first-order classical logic and showed that it decomposes into two main ingredients: a deeper and wider expressive component and, on top of it, a narrower descriptive component. We argued that hypothetical and general claims are fundamental to logic and part of its expressive component. In a sense, our argument is transcendental: these types of claims are what make logic possible. Another face of logic, formed by the alethic and ontic ingredients, flattens - so to speak - the terrain on top of the expressive component giving us the descriptive language of first-order classical logic with identity. It is an important region in the landscape of language, lying among and on top of the larger and original expressive foundation.[25]

[23] As we have noticed, (ONT) is provable if there is only a fixed finite number of objects in the range of 'x'. In this case, the notion of quantification over a closed totality is given automatically by stating that the domain of quantification has the given finite number of elements. For instance, if there are only two objects, we can write '$\exists y \exists z(y \neq z \wedge \forall x(x=y \vee x=z))$'. It is only when the domain is limitless that (ONT) is needed.

[24] Note that a particular case of (ONT) is that $\forall M\,[\forall x(M(x) \vee \neg M(x)) \rightarrow \forall x M(x) \vee \exists x \neg M(x)]$.

[25] Versions of this paper were read at Universidade Nova de Lisboa, Universidade do Porto (on the occasion of the conference *The Logical Alien at 20*) and at

Fernando Ferreira
University of Lisbon

References

Burgess, John. 2005. *Fixing Frege*. Princeton University Press.
Carroll, Lewis. 1895. What the Tortoise Said to Achilles. *Mind* 4: 278-280.
Feferman, Solomon. 1998. Why a Little Bit Goes a Long Way: Logical Foundations of Scientifically Applicable Mathematics. In *In the Light of Logic*. Oxford University Press.
Ferreira, Fernando. 2006. To Catch One's Own Shadow. In *Actas do Segundo Encontro Nacional de Filosofia Analítica*, ed. by Sofia Miguens et al. Faculdade de Letras da Universidade do Porto.
Frege, Gottlob. 1980. Logical Investigations, III: Compound Thoughts. *Posthumous Works*. Blackwell.
Frege, Gottlob. 1984. Logical Investigations, IV: Logical Generality. In *Collected Papers on Mathematics, Logic and Philosophy*. Blackwell.
Kreisel, Georg. 1967. Informal Rigour and Completeness Proofs. In *Problems in the Philosophy of Mathematics*, ed. by Imre Lakatos. North-Holland.
Parsons, Charles. 2008. *Mathematical Thought and its Objects*. Cambridge University Press.
Prawitz, Dag. 1965. *Natural Deduction: a Proof-theoretical Study*. Dover Publications.
Putnam, Hilary. 1998. *Representation and Reality*. The MIT Press.
Quine, Willard. 1953a. On What There Is? In *From a Logical Point of View*. Harvard University Press.
Quine, Willard. 1953b. Two Dogmas of Empiricism. In *From a Logical Point of View*. Harvard University Press.
Quine, Willard. 1976. Truth by Convention. In *The Ways of Paradox and Other Essays*. Harvard University Press. (First published in 1935.)
Russell, Bertrand. 1903. *Principles of Mathematics*. W.W. Norton & Company., 1996.
Russell, Bertrand. 1953. The cult of the 'common usage'. *The British Journal for the Philosophy of Science* 3: 303-307.
Ryle, Gilbert. 1950. 'If', 'so' and 'because'. In *Philosophical Analysis*, ed. by Max Black. Prentice Hall.
Sellars, Wilfrid. 1997. *Empiricism and the Philosophy of Mind*. Harvard University Press.
Simpson, Stephen. 1999. *Subsystems of Second-order Arithmetic*. Springer-Verlag.

Universidade de Lisboa. I would like to thank Nuno Venturinha, Sofia Miguens and João Branquinho for giving me the opportunity to present my work, to Concha Martínez for commenting on it in Oporto and, in general, for the comments and criticism of the audiences. This article was partially supported by the FCT – Fundação para a Ciência e a Tecnologia under the project grant [PTDC/FIL-FCI/109991/2009].

Será o Contratualismo Reconciliável com o Consequencialismo?

Pedro Galvão
Universidade de Lisboa e Grupo LanCog

BIBLID [0873-626X (2012) 34; pp. 625-634]

Será que o contratualismo e o consequencialismo poderão *convergir* na prática? Admitir que sim é julgar que estas perspectivas, apesar de oferecerem fundamentos diferentes para a obrigação moral, poderão implicar exactamente o mesmo quanto à permissividade de todo e qualquer acto, de tal forma que os actos permissíveis de acordo com um contratualista serão permissíveis de acordo com um consequencialista, sendo o mesmo verdade a respeito dos actos impermissíveis.

Este tipo de convergência, vou sustentar, só poderá verificar-se entre versões indefensáveis de contratualismo ou de consequencialismo. Portanto, se alguém argumentar que uma certa perspectiva contratualista converge com uma certa perspectiva consequencialista, teremos de dizer o seguinte: não há realmente convergência entre essas perspectivas ou, se há, pelo menos uma delas será indefensável, o que torna a convergência desinteressante.

O meu argumento contra a possibilidade de uma convergência interessante entre o contratualismo e o consequencialismo baseia-se numa questão acerca da qual vários contratualistas têm divergido efectivamente de vários consequencialistas: a questão das nossas obrigações para com os animais não-humanos. (Por uma questão de concisão, usarei simplesmente o termo «animais» para me referir aos animais das outras espécies.) Vou defender que esta divergência é insanável, argumentando a favor das seguintes teses:

1. Qualquer consequencialismo defensável implica que alguns animais têm estatuto moral.

2. Qualquer contratualismo defensável implica que nenhum animal tem estatuto moral.

Nas próximas duas secções justificarei, respectivamente, 1 e 2. Destas teses segue-se que nenhum consequencialismo defensável converge com um contratualismo defensável.

Afirmar que há animais com estatuto moral, importa esclarecer desde já, é dizer que temos obrigações *para com* alguns animais. Por outras palavras, é pensar que temos deveres *directos* em relação a alguns animais. Se temos estes deveres, um deles será seguramente o de não infligir dor ou sofrimento a animais — concebido apenas como um dever *prima facie*, note-se. Quem considera que este dever é directo entende que não devemos fazer sofrer animais porque o sofrimento é mau *para eles*, vai contra os *seus* interesses, e não simplesmente porque infligir-lhes sofrimento seja mau, de alguma forma, para os seres humanos ou vá contra os nossos interesses. O consequencialista não poderá deixar de aceitar um dever como este; o contratualista, pelo contrário, não poderá admiti-lo.

John C. Harsanyi, embora nunca se tivesse descrito como «contratualista», conta-se entre os autores que argumentaram a favor de uma convergência do contratualismo com o consequencialismo. A última secção deste artigo centra-se no argumento de Harsanyi. Inspeccionando-o, veremos melhor como o esforço de reconciliação parece condenado ao fracasso.

1

Podemos seriar estados de coisas alternativos segundo o seu valor impessoal, adoptando um ponto de vista estritamente imparcial e maximamente abrangente. Directa ou indirectamente, o consequencialista baseia a sua atribuição de permissividade a actos neste tipo de seriação de estados de coisas. No primeiro caso será um consequencialista de actos, sendo de esperar que defenda o seguinte:

> Um acto é permissível se e somente se não há um acto alternativo que resultasse num estado de coisas melhor.

Se optar antes por uma abordagem indirecta, poderá ser um consequencialista de regras e sustentar o seguinte:

> Um acto é permissível se e somente se está em conformidade com um código moral óptimo.
>
> Um código moral é óptimo se e somente se não há um código moral alternativo que resultasse num estado de coisas melhor (se colhesse a aceitação geral).

Há outras versões de consequencialismo, tanto directas como indirectas, mas na presente discussão aquilo que as distingue entre si não é importante.

Alguns consequencialistas são hedonistas. Pensam que *só* o prazer é intrínseca e fundamentalmente bom, e que *só* o sofrimento é intrínseca e fundamentalmente mau. Para estes consequencialistas, dados quaisquer dois estados de coisas alternativos, o melhor será sempre aquele em que houver mais prazer, descontado o sofrimento.

O hedonismo é seguramente uma perspectiva muito controversa e bastante impopular nos tempos que correm. Muitos consequencialistas — a maioria, na verdade — recusar-se-iam a seriar estados de coisas unicamente segundo o critério do «saldo hedónico». Defenderiam que outros critérios — por exemplo, o exercício da autonomia, a satisfação de desejos, a compensação pelo mérito, a igualdade na distribuição do bem-estar ou a integridade dos ecossistemas — são importantes para efectuar as seriações. Como esta lista permite entrever, há muitas questões interessantes a discutir quanto ao modo correcto de avaliar estados de coisas em termos do seu valor impessoal. Mas saber se o sofrimento será *um* dos factores importantes a atender nesse tipo de avaliação não é uma dessas questões. Qualquer consequencialista razoável, por maior que seja a sua aversão ao hedonismo, aceitará o seguinte: *ceteris paribus*, um estado de coisas é melhor do que outro se nele houver menos sofrimento — ou, talvez, menos sofrimento imerecido. Uma perspectiva consequencialista em que os estados de coisas fossem seriados sem se atribuir um peso negativo considerável ao sofrimento resultaria em conclusões práticas extremamente contra-intuitivas, para dizer o mínimo.

Ora, nem só os seres humanos têm a capacidade de sofrer. Ainda que só os membros da nossa espécie estejam sujeitos a certas formas particularmente intensas de sofrimento, partilhamos com muitos outros animais capacidades semelhantes para experienciar dor física, bem como para ter sentimentos de angústia, medo e tédio. De um

ponto de vista estritamente imparcial e maximamente abrangente, o sofrimento dos animais não só não pode ser ignorado como deve ter a mesma importância que o sofrimento semelhante dos seres humanos. Avaliar estados de coisas sem atender ao sofrimento dos membros das outras espécies seria tão *arbitrário* como fazê-lo sem considerar, por exemplo, o sofrimento dos seres humanos de algumas raças ou de um dos sexos.[1]

Admitamos, então, que o consequencialista tem de basear a sua perspectiva numa concepção do valor em que o sofrimento dos animais conte como algo intrinsecamente mau ou desvalioso. Se ele for um consequencialista de actos, incluirá esse sofrimento entre as consequências dos actos que contribuem invariavelmente — ainda que nem sempre de forma decisiva — para os tornar errados. E, se ele for antes um consequencialista de regras, também admitirá um dever *prima facie* de não maltratar animais, já que a aceitação geral de um código moral que sancione este dever teria seguramente melhores consequências do que a aceitação geral de um código moral que consentisse a indiferença total em relação ao bem-estar dos animais.

Qualquer consequencialista, portanto, deverá aprovar uma preocupação directa com os animais, tomando sempre o facto de um acto resultar em sofrimento para um animal como *uma* razão para não o realizar. Não é assim de estranhar que, de Bentham a Singer, os representantes mais destacados da tradição consequencialista tenham figurado entre os defensores de uma maior consideração pelos interesses dos animais. À luz de qualquer versão defensável de consequencialismo, práticas como a criação intensiva de animais, ou o seu uso em laboratórios para fins triviais, deverão ser abolidas em virtude de envolverem uma enorme desproporção entre os males infligidos aos animais e os benefícios que os seres humanos obtêm.

2

Examinemos agora a perspectiva contratualista, que admite a seguinte formulação concisa:

[1] Peter Singer, na sua célebre crítica ao «especismo», defendeu pormenorizadamente que a discriminação baseada na espécie é análoga às discriminações baseadas na raça ou no sexo. Veja-se Singer 1974.

> Um acto é permissível se e somente se está em conformidade com os princípios que os agentes racionais acordariam entre si para governarem a sua interacção.

Os contratualistas baseiam a ética num acordo hipotético, realizado em circunstâncias ideais, mediante o qual os agentes racionais definem os termos da sua coexistência em sociedade. Ora, os animais, mesmo os mais inteligentes, são manifestamente incapazes de participar num acordo — e muito menos num acordo desta natureza, centrado na escolha de princípios morais. Deste modo, parece que as partes contratantes não terão razões para aceitar um código moral em que os interesses dos animais sejam considerados directamente. Peter Carruthers (1992), o contratualista que melhor explorou esta questão, defendeu isto mesmo: que o contratualismo não deixa espaço para a atribuição de estatuto moral aos animais. Os animais, em seu entender, importam apenas indirectamente de um ponto de vista ético. Não devemos tratá-los de certas formas, mas isso apenas devido aos direitos e aos interesses dos seres humanos. Por exemplo, matar certos animais será errado em virtude de violar os direitos dos seus proprietários, e maltratar animais em público poderá ser errado em virtude de ferir a sensibilidade de algumas pessoas.

Há dois tipos principais de contratualismo, que diferem na concepção das circunstâncias ideais do contrato ético.[2] No contratualismo rawlsiano, os agentes desconhecem as suas características e interesses pessoais. Têm conhecimento de factos gerais de natureza psicológica e sociológica, mas, como estão sob um véu de ignorância, ignoram o seu sexo, raça, classe social, bem como os seus talentos naturais e os seus projectos individuais. Por isso, os agentes efectuarão a sua escolha de princípios desconhecendo o lugar que irão ocupar na sociedade (Rawls 1971). No contratualismo scanloniano, pelo contrário, não há qualquer véu de ignorância. Todos os agentes conhecem a sua posição na sociedade e as suas características pessoais. Contudo, partilham o objectivo de encontrar princípios que «ninguém possa rejeitar razoavelmente enquanto base para um acordo geral, livre e informado» (Scanlon 1998: 153).

Consideremos o contratualismo rawlsiano. Dado que os agentes,

[2] Ignoro aqui o contratualismo hobbesiano, cuja incompatibilidade com a atribuição de estatuto moral aos animais é perfeitamente evidente.

sob o véu de ignorância, ignoram a sua raça e o seu sexo, não escolherão princípios que neguem estatuto moral aos seres humanos de algumas raças ou de um dos sexos. Mas, por muito espesso que seja esse véu, os agentes saberão que não se incluem entre os animais, dado que estes são incapazes de participar numa escolha de princípios morais. Deste modo, não admitirão quaisquer obrigações para com os membros das outras espécies.

Uma forma de resistir a este resultado consiste em introduzir a suposição de que, entre as partes contratantes, figuram agentes racionais cuja função é representar os animais e defender os seus interesses. Muito plausivelmente, a participação desses agentes no contrato ético resultaria na escolha de princípios que concedessem estatuto moral aos animais. Esta versão de contratualismo rawlsiano, no entanto, é indefensável: a suposição indicada, sendo destituída de uma justificação independente e servindo apenas conferir estatuto moral aos animais, é puramente *ad hoc* (Carruthers 1992: 99-100).

O contratualismo scanloniano é igualmente adverso à perspectiva de que os animais têm estatuto moral. É verdade, sem dúvida, que muitos dos agentes racionais envolvidos no contrato poderão importar-se com os animais, ter uma preocupação com o seu bem-estar. Mas esta preocupação, como sustenta Carruthers, não oferece uma razão adequada para rejeitar um código moral sem obrigações para com os animais:

> [N]ão será razoável uma pessoa rejeitar uma regra se outras pessoas teriam um fundamento igual para rejeitar *qualquer* regra proposta. Pois, nesse caso, não conseguiríamos atingir o nosso objectivo partilhado de chegar a um acordo geral, livre e sem coacção. Não poderá ser razoável, portanto, rejeitar uma regra simplesmente por esta estar em conflito com um dos meus interesses ou preocupações. Pois qualquer regra (se não for completamente trivial) estará em conflito com as preocupações de alguém. Talvez eu me importe profundamente com o bem-estar dos animais. Mas outros preocupam-se profundamente com padrões de vestuário e aparência, modos de actividade sexual e a adoração do seu deus. Se eu puder rejeitar razoavelmente regras que não atribuem qualquer peso aos interesses dos animais, outros poderão rejeitar de forma igualmente razoável regras que nos permitam vestir e fazer amor como desejarmos, e adorar ou não adorar como bem entendermos. (1992: 104)

Para nos restringirmos ao tema em discussão, suponhamos que o contrato ético envolve, além de agentes que se preocupam com bem

-estar dos animais, agentes que se preocupam de forma igualmente intensa com a manutenção de actividades incompatíveis com a atribuição de estatuto moral aos animais. Numa versão de contratualismo scanloniano em que se considerasse que os primeiros, apelando à sua preocupação, poderiam rejeitar razoavelmente princípios que negassem estatuto moral aos animais, teríamos de admitir que os segundos, apelando à sua preocupação, também poderiam rejeitar razoavelmente princípios que atribuíssem estatuto moral aos animais. E assim, a respeito do tratamento dos animais, não haveria quaisquer princípios que não pudessem ser razoavelmente rejeitados. Esta versão de contratualismo, portanto, seria indefensável: não permitiria que as partes contratantes chegassem alguma vez a um acordo. Sob a perspectiva scanloniana, o facto de algumas pessoas se importarem com bem-estar dos animais não poderá contar como uma razão para escolher princípios que lhes confiram estatuto moral.

Em suma, ao passo que os agentes rawlsianos serão indiferentes aos interesses dos animais, os agentes scanlonianos, ainda que possam importar-se com esses interesses, não poderão apelar à sua preocupação para rejeitar um código moral que os ignore. Num enquadramento contratualista, a perspectiva de que os animais têm estatuto moral não encontra o menor apoio. À luz de qualquer versão defensável de contratualismo, como Carruthers (1992: 107-108, 159-160) sublinha, não há razões para reprovar práticas como a criação intensiva de animais ou o seu uso em laboratórios, mesmo para fins triviais.

3

John Harsanyi defendeu uma ética consequencialista, mais precisamente um utilitarismo de média: *grosso modo*, a perspectiva de que os princípios morais correctos são aqueles cuja adopção maximizaria o bem-estar médio. O seu argumento principal a favor desta perspectiva, embora tenha precedido a teoria da justiça de Rawls, é facilmente enquadrável no contratualismo rawlsiano.[3] Pois Harsanyi, à semelhança de Rawls, alega que os princípios correctos são aqueles que os agentes racionais escolheriam se estivessem sob um véu de igno-

[3] Harsanyi apresentou este argumento em várias ocasiões. Mas veja-se sobretudo Harsanyi 1977.

rância, desconhecendo a posição que irão ocupar no mundo em que os princípios serão implementados. Porém, enquanto Rawls sustenta que os agentes repudiariam princípios utilitaristas, Harsanyi defende o contrário, sugerindo assim uma convergência entre o consequencialismo e o contratualismo — um contratualismo diferente do de Rawls, mas que, em virtude do recurso ao véu de ignorância, não deixa de ser rawlsiano.

O desacordo entre Rawls e Harsanyi respeita sobretudo ao procedimento de escolha apropriado para os agentes racionais. Rawls advoga o *maximin*, um procedimento de escolha extremamente prudente que se deixa traduzir na seguinte perspectiva: os agentes deverão avaliar os princípios como se tivessem a certeza de que vão ficar na posição de um dos indivíduos mais desfavorecidos. Harsanyi sugere antes que os agentes deverão efectuar essa avaliação supondo que a probabilidade de ficarem em cada uma das posições possíveis é rigorosamente a mesma. Ora, estando motivados pelo interesse pessoal, os agentes escolherão assim os princípios que resultem num maior bem-estar médio — em detrimento, por exemplo, dos princípios que resultem num maior bem-estar para os mais desfavorecidos.[4] Deste modo, um contratualismo rawlsiano gera uma perspectiva consequencialista.

Importa agora perguntar se o consequencialismo resultante será defensável. E a resposta, pelas razões indicadas, tem de ser negativa. Dado que os agentes rawlsianos sabem que em caso algum ficarão na posição de um animal, irão ignorar o bem-estar dos animais ao seleccionar os princípios que governarão a sua interacção. Escolherão assim os princípios, se adoptarem o procedimento de escolha que Harsanyi recomenda, cuja adopção resulte num maior bem-estar médio *para os seres humanos*,[5] deixando de fora o bem-estar dos membros de todas as outras espécie. Mas, como vimos, de uma perspectiva consequencialista é arbitrário excluir os interesses dos animais. Portanto, o argumento rawlsiano de Harsanyi conduz a uma versão indefensável de consequencialismo.

[4] Para uma apresentação mais pormenorizada do argumento, veja-se Galvão 2007: 100–102.

[5] Ou, pior ainda, talvez apenas para *alguns* seres humanos, já que nem todos os membros da nossa espécie são agentes racionais.

É verdade que podemos tentar remediar esta falha do argumento. Para que os princípios escolhidos pelos agentes racionais sejam aqueles cuja adopção resultasse num maior bem-estar médio verdadeiramente *geral*, podemos redescrever apropriadamente as circunstâncias ideais em que esses agentes efectuam a escolha. Uma possibilidade, que sugeri noutra ocasião (Galvão 2007: 111-112), consiste em tornar o véu de ignorância um pouco mais espesso, introduzindo uma «condição de amnésia»: antes de efectuarem a escolha, os agentes ficam privados do conhecimento da identidade de todos aqueles cujo bem-estar será afectado. Inicialmente, os agentes poderão saber (1) *quem* são todos os indivíduos afectados — incluindo os animais — e (2) que nível de bem-estar cada indivíduo obterá se os diversos princípios em consideração forem adoptados. Antes de escolherem os princípios, no entanto, os agentes «esquecem» 1, ainda que retenham a informação correspondente a 2. Desta forma, deixam de saber se os indivíduos em questão são animais ou humanos, mas saberão em que medida o seu bem-estar será afectado com a adopção de cada conjunto de princípios. E assim efectuarão a escolha como se pudessem vir a ocupar a posição dos animais, o que os levará a preferir os princípios cuja adopção resulte no maior bem-estar médio, considerados todos os seres sencientes.

Esta nova versão de contratualismo rawlsiano tem, portanto, a vantagem de conduzir a um consequencialismo que não deixa de fora os interesses dos animais — e que por isso seria indefensável. Contudo, a «condição da amnésia» afigura-se puramente *ad hoc*, pelo torna indefensável qualquer perspectiva que a inclua.

Em suma, o argumento inicial de Harsanyi gera um consequencialismo indefensável. Ao repararmos a sua falha, acabamos com um contratualismo indefensável. Depois da discussão realizada nas duas secções anteriores, este resultado não deve surpreender-nos: se qualquer consequencialismo defensável diverge de qualquer contratualismo defensável no que respeita ao estatuto moral dos animais, a convergência entre estas posições só se realiza quando pelo menos uma delas se apresenta numa versão manifestamente indefensável. Se o argumento aqui apresentado é sólido, a questão de saber se haveremos de ser consequencialistas ou contratualistas mantém, pois, uma

importância prática muito significativa.[6]

Pedro Galvão
Universidade de Lisboa e Grupo LanCog

Referências

Carruthers, Peter. 1992. *The Animals Issue: Moral Theory in Practice*. Cambridge: Cambridge University Press.
Galvão, Pedro. 2007. *Do Ponto de Vista do Universo: Um Estudo sobre a Racionalidade da Ética Consequencialista*. Lisboa: Centro de Filosofia da Universidade de Lisboa.
Harsanyi, John C. 1977. Morality and the Theory of Rational Behaviour. In A. Sen e B. Williams. *Utilitarianism and Beyond*. Cambridge: Cambridge University Press, 1982, 39-62.
Rawls, John. 1971. *A Theory of Justice*. Cambridge: Harvard University Press.
Scanlon, T. M. 1998. *What We Owe to Each Other*. Cambridge: Harvard University Press.
Singer, Peter. 1974. All Animals Are Equal. *Philosophical Exchange*, 1: 103-116.

[6] Através de uma Bolsa de Pós-Doutoramento (SFRH/BPD/27852/ 2006), beneficiei, na realização deste trabalho, do apoio da Fundação para a Ciência e a Tecnologia no âmbito do III Quadro Comunitário de Apoio.

Este artigo *não* obedece ao Acordo Ortográfico de 1990.

Physicalism and Early Behaviourism

Nelson Gonçalves Gomes
University of Brasília

BIBL[0873-626X (2012) 34. pp. 635-643]

In the history of philosophy very few thinkers opposed metaphysics in such a radical way as Otto Neurath (1882-1945), a member of the so called First Vienna Circle (1907-12), and an active participant in the Moritz Schlick's *privatissimum* that regularly met for fifteen years from October 1923 onwards. Neurath was a fertile thinker, as he wrote over five hundred different works comprehending areas like philosophy, economics, sociology, and education. He was also a restless social organiser and editor of scientific publications. It is true that mostly of his essays are rather short and full of repetitions, but his thought was in continuous transformation from his very first work on (1909). In the present article we will concentrate on the ideas developed by Neurath on physicalism in the late 1920 and in the first half of the 1930s. In that period Schlick's *privatissimum* was termed Vienna Circle thank to the publication of the well known *manifesto*, which was actually written by Neurath with some opinion by other fellow-members especially Carnap. Philosophically the *manifesto* was a compromise between some intellectual tendencies within the Circle, which is a reason why it is not the best reference for someone who wants to understand Neurath's anti-philosophy. In these lines we will present the Neurathian view of physicalism. Our effort will be to show that early behaviourist psychology was introduced in it as a crucial element to complete what Neurath believed to be a radical version of materialism.

In a review of Carnap's *Aufbau* written in a political newspaper in 1928 Neurath speaks favourably about the anti-metaphysical tendency of the book but he clearly expresses his objections relative to the limits of some Carnapian claims to sharp logical order and theoreti-

cal exactness. (Neurath 1981/1928: 295-7) Accordingly to Neurath a sociologist knows that ambiguity is unavoidable to a certain extent, which gives rise to problems that only explicit ideological choices may solve. He was indeed a materialist deeply engaged in the project of a Marxian sort of society. Science for him is a revolutionary power, capable of forecasting what is going to happen. This wonderful ability of science makes it possible for mankind to find reliable guidance about individual and especially about collective decisions. Philosophy in general and metaphysics in special are for him empty and dangerous forms of talk produced by enemies of the working class. As Neurath repeats *ad nauseam*, an essential task of the Vienna Circle is the full elimination of metaphysics and philosophy, which entails a thorough cleansing of all philosophic-speculative remains within scientific language. (To quote only a few examples: Neurath 1981/29B, 1981/31A, B, C and D, 1981/1932A, 1981/1933A) About the language of physics Neurath is optimistic for he believes its received metaphysical nonsense is cleared away for good. But he is rather critical about languages of sociology and psychology that wrap up a great deal of speculative pseudo-concepts. In sociology Neurath's main target is Max Weber whose social thought is taken by him as a privileged example of an unacceptable view of society. (Neurath, 1981/1931B: 415 and D: 462-5) Speaking about his contemporary psychology, Neurath considers Watson's behaviourism, the gestalt, Freudian psychoanalysis and individual therapy. He has minor criticisms against psychoanalysis, but he is rather stern about the other tendencies. Watson's discussions on ethics are plainly dropped as metaphysics. (Neurath 1981/1933: 601-8)

In 1931 Neurath popularises his views on physicalism in two short but enlightening articles. One of them was published in English in the USA (Neurath, 1981/1931B), while the other was written in German and published in Italy. (Neurath, 1981/1931C) In both of these works, besides his usual criticisms against metaphysics and philosophy, he explains very clearly his theses on language and unified science. For him, language is present from the very beginning in any sort of intelligent activity. Therefore, it would be meaningless to try to validate language through pre-linguistic resources like comparing it with the world, in order to make sure that it is really adequate. Even a simple language is about bodies, their properties and rela-

tions, which expands to every meaningful kind of human linguistic expression. Science develops within language and cannot be severed from it. But the alpha and the omega of science are its forecasts. The business of science has nothing to do with any sort of description of the world nor with speculation, for science is public knowledge, whose sentences are due to be submitted to public tests. Eventually science produces forecasts useful to society in its activities of deciding and planning.

In these 1931 articles Neurath emphasizes the importance of behavioural psychology, despite his rejection of its specific theses. Almost two decades before, in 1913, John Watson had published in the USA his well-known *manifesto* termed *Psychology as the Behaviorist Views It*. (Watson, 1913) Watson's programme was to build a psychology concerned with observable behaviour, in order to put it on the path of sciences like physics, chemistry and biology. But Neurath surprisingly inverts this order, saying that sciences are concerned with the behaviour of nature: mineralogy for instance studies the behaviour of stones. (Neurath, 1981/31B: 413-4). In this sense every science is behavioural, because it deals with the observable conduct of objects that are in space and time. Watson was engaged in a form of psychology that, generally speaking, gets rid of concepts like mind, consciousness and the like. Introspection in particular is strongly rejected by Watsonian psychology. This is precisely the frame that Neurath includes in his version of physicalism, especially in relation to sociology. In a book published also in 1931, he describes sociology as 'social physics' (an expression originally by Quételet) or 'social behaviourism'. (Neurath 1981/31D: 500-26) Sociology has nothing to do with empty ideas like *Volksgeist* for instance, which are popular among social scientists, despite their character of metaphysical remains waiting to be cleared away. The true business of sociology is the study of human beings, who are stimulated by other human beings and by their environment. They work, they play, they are consumers, or, briefly, they behave.

Neurath extends this behavioural view to the statements of an individual person. If someone says, for example, 'I see something blue' this statement will be taken as an empirical sentence (*Realsatz*) if outside of the person's body some space-temporal things of the correspondent kind happen. The statement will be considered to be

a hallucination if everything that happens is some change in the person's brain. Finally the statement will be taken as a lie if only the person's language-centre and its expression are changed. Thank to this type of scheme, Neurath is able of comprehend human expressions in purely behavioural terms. (Neurath 1981/1931C: 420)

In his two popular articles of 1931 Neurath presents the process of constitution of the unified science (*Einheitswissenschaft*), which is the sequence of space-temporal sentences together with their ramifications. The unified science includes logics and its tautologies, which are only rules for the transformation of sentences. Proper science starts with observational statements (*Beobachtungssätze*). Induction has no theoretical justification, but it is actually used based upon some methodological decision. Scientific laws obtain inductively from singular sentences. Such laws are not proper universal sentences, but they are only rules for the inference of statements related to singular forecasts, which have to be submitted to empirical tests. Neurath emphatically sustains the thesis that sentences cannot be compared with the world, but only with other sentences. Without any detail, he says that sentences are submitted to empirical tests within contexts where light, sounds and the like are present. Despite this, light and sounds are not elements of the unified science, whose only constituents are sentences. The test of a sentence produces other sentences. If a single sentence can be consistently added to the unified science, it is considered to be "true." If not, it is taken as "false." It is not comfortable for Neurath to use philosophical language, which is the reason why he writes words like these with inverted commas. For him, there is no other way of speaking about "truth." (Neurath 1981/1931C: 418-9) He refused the accusation of being a coherentist.

In 1931 Carnap publishes an article on the physicalist language as the universal language of science, where he partially preserves some phenomenalist elements of the *Aufbau*. The test of sentences of an empirical system, Carnap says, leads eventually to protocol or observation statements (*Protokollsätze*), which deal directly with basic experiences. Such elementary statements need no further tests. (Carnap 1931: 437-41)

In German "Protokoll" means something like a register or record. In this sense, protocol-statements are reliable registers of what a sci-

entist verifies within a single experiment, under particular space-temporal conditions. For instance, if he describes what happened in his experiment the scientist writes down records of a day, of an hour, of a place and other conditions. Eventually he may write something like "A noise was heard followed by a smell of ozone". This sentence is phenomenic, because it describes what was individually perceived by the scientist. Accordingly to Carnap, the test is complete when protocol-statements obtain.

A strong behaviourist and physicalist like Neurath could not accept Carnap's thesis, which caused him to publish one of his most elaborate and artful works: his essay on protocol-statements, originally published in the journal *Erkenntnis*, in1932. (Neurath 1981/1932B) In this article Neurath reaffirms his idea of a physicalist language as something that human beings have from their early life, as trivial historic language is taught to everyone since childhood. On the other extremity, sophisticated scientific language include highly complex space-temporal co-ordinates. Neurath persists in his programme towards the actual constitution of this kind of physicalist language, which for him would be the only possible task for the Vienna Circle. The programme should firstly involve a complete cleansing of all metaphysical remains within natural language. After this Neurath's language for the unified science would be woven like a seamless piece of cloth. As Neurath remarks, trivial colloquial language allows nonsensical expressions like 'a neighbour without a neighbour' for instance, which may be the root of speculative pseudo-problems. But the language for unified science is not supposed to be absolutely clean and transparent, because it is impossible to eliminate all sorts of ambiguities in it. Such a language cannot get rid of the so-called "agglomerations" (*Ballungen*), which are imprecise terms that cannot be replaced by sharply defined ones. (Neurath 1981/1935) If an "agglomeration" is eliminated somewhere, it will reappear in another point. As a matter of fact, Neurath vividly opposed any idea of a flawless language. But imprecision is not a logical mistake.

In addition to this, Neurath also opposed any sort of metalanguage for his unified language, because one part of the latter may speak about another of its parts. The unified language does not need a second language to deal with it.

The original point in Neurath's main article on protocol-state-

ments is his formulation of what is a protocol. He drops his former speech about observational sentences in favour of Carnapian protocols, but his meaning is totally different. Neurath illustrates his idea with a famous example:

> "Otto's protocol at 9:17 a.m.: [At 9:16 a.m. Otto's linguistic thought was like this: (At 9:15 a.m. there was in the room one table perceived by Otto.)]" (Neurath 1981/1932B: 580)

By elimination of square and round parentheses in this example, empirical sentences obtain, which are not protocols. A protocol must contain a person's name and a term relative to perception.

At first sight this characterisation looks rather weird, but it tries to capture the idea of a register written by someone who identifies himself and his perception. Nevertheless Neurath does not intend to admit irreducible psychological concepts. On the contrary: he insists that everything in a protocol-statement may be translatable into a space-temporal language, i.e. into a behavioural language. The important item is that the name and the perception-term indicate reference-points for the test of a protocol. If changes outside Otto's body happen, his statement will be empirical; if changes are limited to areas of his brain, the statement will be a hallucination; if only his language-centre is involved, the statement will be a lie. Anyway a protocol-statement has to be submitted to tests. Independently of this, and contrarily to Carnap's thesis, a protocol may be dropped, whenever it fails to be consistently included in the unified science.

Under Neurath's view there is one physicalist language and one unified science where protocol-statements are included. In science no sentence is privileged or untouchable. Any sentence is dischargeable and so is the set of sentences where they are supposed to be included. If there is a contradiction between a protocol-statement of mine and a protocol by someone else, I may prefer the latter to the former. What is not admissible is a contradiction within the set of scientific sentences. If this happens, it is theoretically possible to keep one protocol-statement and to drop the set, which we hardly ever do. In science nothing is immune to revision.

Each individual person may test protocol-statements formulated by others, and he may do the same with his own protocols, whereby the procedures are the same. Unified science is social and it is

changing all the time. Robinson Crusoe and Friday share the same language, a reason why they can mutually test their protocol-statements. In this context, Neurath repeats his famous metaphor, that we are like sailors, forced to rebuild our ships in high sea. Science is unstable but crucial for us, because of its ability to forecast. If an alien protocol-statement is useful to a better prediction, we may prefer it to our own formulations.

The idea of behaviourism Neurath borrows from Watson clearly plays an essential role as far as it makes possible, at least in principle, the reduction of all human conduct to space-temporal elements, which eliminates the Carnapian phenomenalism and even the distinction between two kinds of psychical contexts present in the *Aufbau*. Thanks to the overall behavioural frame, physicalism can avoid talking about mind, consciousness, introspection and the like. This eliminates traditional psychology and allows Neurath's programme to be complete. Scientific sentences and protocols are not submitted to tests by means of procedures that lead to what one meets only in his own self. On the contrary: within the context of the same language, a test involves resources to the same physical and behavioural course of action. Behaviourism is the touch-stone of Neurathian physicalism.

Surprisingly enough Neurath speaks rather little about Watson's behaviourism, whose most general attitude he accepts, but whose specific theories he considered vague or even nonsensical. In order to distance himself from that early behaviourism, Neurath proposed new German words: *Behaviorist* is a representative of Watson's behaviourism; *Behavioristiker* is someone committed with the programme of a physicalistic language for the unified science. This programme extends very naturally into a social behaviourism (*Sozialbehavioristik*), that Neurath brings near to Marxism. (Neurath 1981/1933: 600-4)

It is true that Neurath speaks about behaviourism and about physicalism as programmes or tasks to be developed, because they actually do not exist yet. His characterisation of the language of unified science is a remarkable effort that ought not to conceal a considerable degree of imprecision and even of arbitrariness. For instance Neurath's thesis about the relation (or lack of relation) between sentences and the world is not explained in any detail by him. His concern to rule out metaphysics takes up an important part of his writings, in

disfavour of some interesting contributions he gives. It is true that Neurath plainly did not want to make any sort of philosophy, but this does not justify any theoretical arbitrariness. He was always prepared to apply a sharp Neurathian razor to every statement he considered to be metaphysical, without establishing criteria for this sort of judgement.

In contemporary terms, Neurath can be classified as a naturalist and anti-realist. Neurath is a naturalist because he doesn't admit the existence of any level of knowledge superior to science, able to evaluate science itself. Anti-realist because his view of science is that of the boat to be kept afloat, insofar as it forecasts future events, which has nothing to do with any description of reality. In other words, the set of scientific sentences is a socio-linguistic fact, whose nature is pragmatic.

Neurath's pragmatism is very present in his objections against metalanguage and semantics. For him both are metaphysics. His theses on the subject and his controversies with other physicalists are by themselves a chapter in the history of the later Vienna Circle. Independently of this and within a different context, Michael Dummett writes in 2010, that a semantic theory has a link with metaphysics. (Dummett 2010: 130) If Dummett is right, Neurath's criticisms against semantics are not based on misunderstanding, at least from the point of view of his particular brand of physicalism.

Neurath has been classified as a post-modernist as well, for he represents the view that the promise of Enlightenment has to be reconceptualised. (Cartwright *et al.* 1996: 254) Naturalist, anti-realist, post-modernist: all these descriptions prove that Neurath's thought is a challenge, just because it is so comprehensive and radical.

Nelson Gomes
University of Brasília

References

Carnap, Rudolf. 1931. Die physikalische Sprache als Universalsprache der Wissenschaft. *Erkenntnis*, **2**: 432-65.
Cartwright, Nancy *et al.* 1996. *Otto Neurath - Philosophy Between Science and Politics*. New York: Cambridge University Press.
Dummett, Michael. 2010. *The Nature and Future of Philosophy*. New York: Colum-

bia University Press.
Neurath, Otto. 1981/1929A. Wissenschaftliche Weltauffassung - Der Wiener Kreis. Neurath, Otto. 1981, pp. 299-336.
Neurath, Otto. 1981/1929B. Wissenschaftliche Weltauffassung. Neurath, Otto. 1981, pp. 345-8.
Neurath, Otto. 1981/1931A. Weltanschauung und Marxismus. Neurath, Otto. 1981, pp. 407-12.
Neurath, Otto. 1981/1931B. Physikalismus - Die Philosophie des Wiener Kreises. Neurath, Otto. 1981, pp. 413-6.
Neurath, Otto. 1981/1931C. Physikalismus. Neurath, Otto. 1981, pp. 417-22.
Neurath, Otto. 1981/1931D. Empirische Soziologie. Neurath, Otto. 1981, pp. 423-528.
Neurath, Otto. 1981/1932A. Die „Philosophie" im Kampf gegen den Fortschritt der Wissenschaft. Neurath, Otto. 1981, pp. 571-76.
Neurath, Otto. 1981/1932B. Protokollsätze. Neurath, Otto. 1981, pp. 577-86.
Neurath, Otto. 1981/1933. Einheitswissenschaft und Psychologie. Neurath, Otto. 1981, pp. 587-610.
Neurath, Otto. 1981/1935. Einheit der Wissenschaft als Aufgabe. Neurath, Otto. 1981, pp. 625-630.
Neurath, Otto. 1981. *Gesammelte philosophische und methodologische Schriften*. 2 vols. ed. by Rudolf Haller und Reiner Rutte. Vienna: Hölder-Pichler-Tempsky.
Watson, John. 1913. Psychology as the Behaviorist Views It. *Psychological Review*, **20**: 158-77.

A Typology of Conceptual Explications

Dirk Greimann
Fluminense Federal University

BIBLID [0873-626X (2012) 34; pp. 645-670]

Conceptual explication is an art having approximately the same importance for philosophy that logical deduction has for mathematics. To be sure, conceptual explication is important, not only for philosophy, but also for any other science. For all sciences are committed to clarify their own concepts. But philosophy is tied to conceptual explication in a special way, because the philosopher's work consists in large part in explicating concepts, just as the mathematician's work consists in large part in deriving theorems. For this reason, philosophy has closer ties with conceptual explication than does any other science.

With regard to logical deduction, we already have elaborate theories at our disposal that explain the rules we follow implicitly when we are making logical deductions. Unfortunately, this does not apply also to the explication of concepts. On the contrary, it seems that Carnap's diagnosis of the state of the art from 1950 is still valid, according to which "(p)hilosophers, scientists and mathematicians make explications very frequently", but "they do not often discuss the general rules which they follow implicitly" (Carnap 1950, p. 7).

In what follows, my aim is to make some steps toward the construction of a theory of conceptual explication whose tasks are basically the following four: (i) the explication of "explication"; (ii) the distinction of the main types of explication; (iii) the formulation and justification of criteria of adequacy for each type; (iv) the construction of a system of rules for each type whose observation guarantees the adequacy of the corresponding conceptual explications. Such a theory would not only contribute to the methodological self-awareness of philosophy, but it may also help to resolve some of

the confusions that have already crept into the practice of conceptual explication. Thus, Richard Kirkham has persuasively shown that the current debate about the concept of truth suffers from a "multidimensional" confusion that results from mixing up different projects of explication with different goals and different criteria of adequacy.[1] The clear distinction of these projects and the corresponding types of explication is perhaps the most important desideratum of the current theory of truth.

For this reason, I shall focus in this paper on the distinction of the various types of conceptual explications.[2] The main goal is to set up a typology (or "taxonomy") of conceptual explications. The paper is structured as follows. In section 1, the concept of "conceptual explication" is briefly explicated. Section 2 is devoted to the task of explaining the criteria of classification for conceptual explications on which the typology is based. In sections 3 to 7, the main types of conceptual explication are distinguished and the corresponding criteria of adequacy are sketched. Finally, in section 8, the adequacy of the proposed typology is briefly discussed.

1 Explication of "explication"

In the history of philosophy, conceptual explications have been made partly with a descriptive and partly with a revisionary intention.[3] The aim of a descriptive explication is to describe the content and function of a given concept as faithfully as possible. A revisionary explication, on the other hand, aims to transform a concept that is useful but somehow defective from a scientific point of view into a scientifically more respectable concept. Such defects may be, e.g., the vagueness of a given concept, its inconsistency or the lack of a proper principle of individuation for it. The linguistic theories of truth such as the disquotation theory and the minimalist theory are examples of descriptive explications, and the epistemological theo-

[1] Cf. Kirkham 1995, chapter 1.

[2] In Geimann (2007), I concentrate on the methodological rules for the correct explication of concepts.

[3] This distinction is analogous to Strawson's well-known distinction between descriptive and revisionary metaphysics.

ries like the consensus and the verification theory are examples of revisionary ones.

Following Quine, it has become common practice in large parts of Analytic Philosophy to call descriptive explications "conceptual analyses" and revisionary explications simply "conceptual explications".[4] This terminology may be legitimate, but it is certainly not appropriate with regard to our ends. Since we want to construct a classification of conceptual explications, we need a more general notion of explication that allows us to consider descriptive and revisionary explications as special cases. To take this into account, I shall follow here the traditional approach to conceive of an explication as an answer to a question of the form, "What is X?" like "What is truth?" and "What is knowledge?" A descriptive explication of truth, for instance, explains what is actually understood by "truth" in natural language, and a revisionary one what should be understood by "truth" in science.

According to Quine, the essence of conceptual explication consists in the elimination of the explicandum, that is, its substitution by a similar but scientifically more respectable concept. He in fact defines: *explicare est eliminare* (1960, § 53). Even when we restrict this explication of "explication" to the revisionary case, it does not appear to be satisfactory. For, in order to transform a defective concept into a scientifically more respectable one, it is neither possible nor necessary in all cases to replace it by another one. Take, for instance, the pretheoretic concept of set. From the point of view of set theory, this concept is vague and ambiguous. But, to overcome these defects, it is not possible to substitute the explicandum by another concept, because the concept of set is primitive. Nor is this necessary, because the defects can be overcome by means of an axiomatic characterization or kindred methods.

Moreover, Quine's explication of "explication" does not seem to be materially adequate, for two reasons. First, the explicandum and the explicans are not sufficiently similar: to "explicate" does not mean to "eliminate", but to make something explicit. Second, this explication does not take into account that the majority of the philosophical concepts are "primitive". It is true that in set theory and

[4] Cf. Quine 1960, §§ 39, 53.

other formal disciplines the conceptual explications actually given consist in the elimination of the explicanda, that is, more precisely, in their reduction to a few basic concepts. But the same does not apply to philosophy, because the main philosophical concepts (such as existence, truth, meaning, freedom and morality) cannot be reduced to more fundamental concepts.[5] The elimination of the explicanda is hence a wrong ideal in philosophy.

What we have called here "conceptual explication" is frequently called "definition" in the literature.[6] Unfortunately, the term "definition" is extremely vague and ill-defined. A definition in Frege's sense consists in the logical decomposition of a complex sense into its constituents or in a stipulation to the effect that the definiendum has the same sense as the definiendum.[7] His thesis of the indefinability of truth says that truth is a logically simple (atomic) concept. A definition in Tarski's sense is something quite different. To define X in his sense means to formulate an open sentence whose extension is X, where X may be an individual or a set. His theorem of the indefinability of truth says that in theories whose language is semantically closed it is impossible to construct in a consistent way an open sentence whose extension is the set of all true sentences. To say that truth is indefinable in this sense is to say that truth is inexpressible.[8] Other authors mean by "definition" the reduction of the definiens to more fundamental concepts. When, for instance, Davidson speaks of the "folly of trying to define truth", what he has in mind is the reduction of truth to more fundamental concepts.[9] In order not to complicate things unnecessarily, I shall not try to define "definition" here, but to avoid using this term.

[5] This is confirmed by the fact that no philosophical concept that has been successfully reduced to other concepts until now.

[6] See, for instance, Gupta 2008.

[7] Cf. Frege 1914, pp. 226-228.

[8] Cf. Tarski 1933.

[9] Cf. Davidson 1996.

2 Criteria for distinguishing types of conceptual explication

When we want to construe a typology of conceptual explications, we need criteria of classification that enable us to distinguish between different types of conceptual explications. Such criteria may refer either to the purpose of an explication, or to the method used to realize the explication, or to the kind of its explicandum, or to its criteria of adequacy.

In the literature it is common to distinguish between three basic types of explications with regard to their criteria of material adequacy: descriptive, revisionary and stipulative explications. A descriptive explication is materially adequate only if it is faithful to the original meaning of the explicandum, i.e., the explicandum and the explicans must have the same meaning. In the case of revisionary explications, by contrast, the explicans and the explicandum are not supposed to have the same meaning, because otherwise the explicans could not be transformed into a scientifically more respectable concept; you cannot wash the fur without making it wet. Nevertheless, in the revisionary case the explicandum and the explicans are supposed to have similar meanings, because otherwise the explicans could not substitute the explicandum. Finally, in the case of stipulative explications, the explicandum and the explicans need not even have similar meanings, because everyone is free to stipulate what he wants to understand by the explicandum.

A second classification refers to the method used in explications. In this respect, we may distinguish between explicit and implicit definitions, ostensive and lexical definitions, operative and contextual definitions, axiomatic characterizations, Fregean elucidations, ramseyfications etc.

The third criterion of classification refers to the kind of the object of explication, that is, the type of the explicandum. In this respect we may distinguish between extensional and intensional explications, and between nominal and real explications.

Finally, the fourth criterion of classification refers to the purpose of explications. Thus, descriptive and revisionary explications differ, in the first place, with regard to their purposes. The same applies to the distinction between stipulative and descriptive explications.

It is clear that among the four criteria mentioned the last two are the most important ones. For, the methods of explications and their criteria of adequacy depend on their objects and their purposes. It can, moreover, be shown that a larger part of the confusions to be found in the literature about the explication of truth refer to the objects and the purposes of explications. For these reasons, I shall concentrate in what follows on the classification of explications with regard to their objects and their purposes.

3 Real and nominal explications

When we want to answer the question "What is water?", we may consult either a linguist or a chemist. The linguist will explicate the linguistic meaning of "water", that is, the criteria of the correct application of the word "water". Part of his explication may be that "water" applies to those entities that are liquid, colourless, odourless, can be used to satisfy one's thirst, and so on.[10] The explicandum is in this case the descriptive content of "water", its Fregean sense. When, on the other hand, the chemist answers the question "What is water", he does not explicate the semantic structure of "water", but the chemical structure of water.

What is traditionally called a "nominal definition" is an explication whose object is a given term ("nomen"). By contrast, the object of a "real definition" is the thing ("res") denoted by that term. More precisely, the object of a nominal explication is the *meaning* of a given term, and the object of a real explication the *nature* or *essence* of the thing denoted by that term. The lexical explications of "water" to be found in dictionaries are examples of nominal explications, and the definition of water to be found in chemistry textbooks, "Water is H_2O", is an example of a real explication.

In the current debate about truth, the distinction between nominal and real explications is largely ignored. The reason probably is that, for most authors, this distinction is a "distinction without a

[10] I am presupposing here that "x is water" means "x is something having the same phenotype as water". In some contexts, "x is water" may mean "x is something having the same microphysical (chemical) structure as water (on earth)". For simplicity's sake, I shall ignore this problem.

difference". They identify the essence of a thing with the sense of the corresponding term. An exception is William Alston, who accounts for the distinction in an indirect way, by distinguishing between concepts and properties. In his seminal article "Truth: Concept and Property", he argues that "a property might have various features not reflected in our concept of that property" (2002, p. 12). Thus, the property of heat is revealed by physics to be the average kinetic energy of constituent molecules, even though our ordinary, pretheoretic concept of heat involves no such component. This concept consists of the set of features that any competent speaker associates with the word "heat", the "criteria" all speakers use when they apply that term. Obviously, ordinary speakers do not associate "heat" with the average kinetic energy of constituent molecules. For this reason, this feature does not belong to the ordinary concept of heat, though it may be a feature of the corresponding property.

According to Alston, the same applies to the concept and the property of truth: the property involves features that are not reflected in the concept. He argues that the minimalist account of truth, according to which the concept of truth is exhaustively explicated by the instances of Tarski's truth-scheme, is adequate when it is considered to be an explication of the concept of truth. The features of the property of truth that are not contained in the concept of truth and that are not captured by the minimalist account are made explicit by the correspondence theory of truth. Examples are the features that a true sentence is a sentence that corresponds to a fact and that true sentences are made true by facts. Just as the explication "Water is H_2O" correctly explicates the property of being water, but not the concept of water, so too the explication "A true sentence is a sentence that corresponds to a fact" explicates correctly the property of truth, but not the concept of truth.[11]

Frege's theory of sense and reference may also be regarded as an explication of the property of being true. According to it, the truth of a sentence depends on the sense and the reference of the words that occur in it. An elementary sentence of the form "a is F" is true

[11] This means, in Kantian terms, that this explication is not an analytical sentence, but a synthetic one. To arrive at it, it does not suffice to analyze concepts; it is necessary to analyze the structure of the world.

if and only if the object determined by the sense of "a" possesses the property determined by the sense of "is F". This explication is not meant as an explication of our concept of truth, which Frege considers to be primitive, but as an analysis of the semantic factors on which the truth of a sentence depends.

Elaborating Alston's distinction, we could say that concepts are constituents of thoughts and properties are constituents of facts. Just as two different senses expressed by proper names (such as the senses of "the evening star" and "the morning star") may refer to the same object, so too two different senses expressed by two one-place predicates (concepts) may refer to the same property, and two different thoughts may refer to the same fact.

A second possibility to account for the distinction between nominal and real definitions is based on the distinction between two senses of identity sentences such as "Water is H_2O", "Pain is C-fibre firing" and "Truth is correspondence with a fact". On the first reading, "is" means identity and on the second reading constitution.[12] Suppose, for instance, that properties are identical when the corresponding concepts are identical, that is, when the corresponding terms have the same Fregean meaning, the same criteria of application. In this case being water and being H_2O are different properties. Even in this case we can maintain that water is H_2O, in the sense that the property of being water is ontologically constituted (or "realized") by the property of being H_2O. Similarly, we may defend the thesis that truth is the correspondence with a fact, even when we admit that truth and correspondence with a fact are different properties.

On this approach, a real explication of a given property aims to make its ontological structure explicit – how the property is constituted. A nominal explication, on the other hand, seeks to make its semantic structure of the explicit – what the features are of which it is composed.

The distinction between nominal and real explications is important for at least three reasons. First, it shows that conceptual explications that are commonly considered to be incompatible might actually be complementary. Thus, if Alston is right, the minimalist

[12] This paragraph is inspired by the distinctions made in Brink 1989, pp. 156-163.

account of truth can be made compatible with the correspondence theory when the first is construed as an explication of the concept of truth and the second as an explication of the property of truth (in his sense of "concept" and "property").

Second, the distinction is important for the evaluation of Moorean arguments against some projects of conceptual explication.[13] To see this, consider the following example from metaethics. According to Moore's argument, a definition of morality in terms of "natural" terms like "A morally right action is an action maximizing human welfare" cannot be materially adequate, because the definiens and the definiendum have the same meaning. His argument is that sentence-pairs like "A morally right action is a morally right action" and "A morally right action is an action that maximizes human welfare" have different meanings: whereas the first sentence is analytic and trivial, the second is synthetic and informative.

This argument implies that the sentence "A morally right action is an action that maximizes human welfare" cannot be considered as an adequate explication of the *meaning* of "is a morally right action". But this does not rule out that it correctly explains the *reference* of "is a morally right action", what moral rightness is and in what it consists. The argument implies, in other words, that the project of giving a nominal explication of morality must be abandoned, but it does not imply that the project of giving a real explication of morality must also be abandoned.

Thirdly, the distinction between nominal and real explications is important for the discussion of deflationist arguments against philosophical theories of truth and non-cognitivist arguments against normative ethics. It is commonly assumed that a radically deflationist account of truth like the redundancy theory has devastating implications for those accounts which consider truth to be a property. For, if sentence pairs like "Snow is white" and "The thought that snow is white is true" express the same meaning, then there is no property of truth.[14] This consequence seems to imply that the endeavour of philosophical theories of truth to uncover the nature of the property of truth is doomed to failure right from the start. Similarly, the

[13] For more details, see Brink 1989, Chapter 6.

[14] This consequence is drawn, for instance, in Brandom 1998.

non-cognitivist accounts of morality, according to which "is morally right" is not used to ascribe a property, but to indicate the performance of an expressive or directive speech act, seems to undermine the possibility of normative ethics.

In fact, however, the consequences of the deflationist accounts of truth and the non-cognitivist accounts of morality are less dramatic, because the meanings of terms in natural language are relevant only for the adequacy of nominal explications, and not also of real explications. Suppose, for instance, that the redundancy account of truth is correct. In this case, there is no property to which the term "true" refers. But this does not imply that the property analyzed by the correspondence theory does not exist. If the correspondence theorist considers his account of truth as a real explication, he can argue that he is not interested in the semantics of "true", but only in the property to which he refers in his theory by means of "true". He may admit that, in natural language, "true" does not refer to any property, and to take this into account, he may replace this word in his theory by a similar one, say "frue", as Tarski suggested in a parallel case.[15] Natural language does not decide which properties exist. Similarly, the non-cognitivist account of morality does not imply that there are no ethical properties, but only, that these properties are not denoted by the corresponding terms of naturally language.

The distinction between intensional and extensional explications also refers to the objects of explications. In the case of intensional explications, the object is the meaning of the explicandum, and in the case of extensional explications, its extension. An example of an extensional explication is Tarski's definition of truth; its goal is to define the set of true sentences of the language of the class calculus.[16] His theorem of the undefinability of truth does not say that the meaning of the truth-predicate or the features of the property denoted by it cannot be made explicit, but only that in a semantically closed language it is impossible to construe in a consistent way an open sentence whose extension is the set of true sentences of that language.

We have seen that an important criterion of classification for

[15] Cf. Tarski 1944, p. 356.

[16] Cf. the definition of truth in Tarski 1933.

conceptual explications concerns the type of the explicandum. The explicandum can be either the meaning of a given term, or the property to which it refers, or the extension of this property. A second important criterion of classification concerns the purposes for which explications serve. In this regard, we must distinguish between descriptive, revisionary, reductive and stipulative explications.

4 Descriptive explications

A major part of the explications to be found in the history of philosophy are devoted to the task to make the content of a given concept explicit and to distinguish it from other concepts with which it is easily confused. Such explications may be called "descriptive" explications. The larger philosophical programs into which they are embedded pursue either a propaedeutic or a therapeutic goal. The descriptive explications can accordingly be subdivided into "propaedeutic" and "therapeutic" ones.

4.1 Propaedeutic explications

The context in which propaedeutic explications are typically made are the programs of the "metaphilosophical" disciplines such as meta-ethics, meta-logic, meta-metaphysics, meta-mathematics and so on. Their purpose is to prepare the construction of the corresponding philosophical discipline by making its central concepts clear and distinct.

Kant's explication of moral goodness is an example of an explication of this type. He distinguishes between good and evil on the one hand, and well-being and ill-being on the other, and then goes on to show that these pairs of concepts, which are easily confused, are independent of each other: it is not a contradiction that an evil man is doing well. According to Kant, this conceptual distinction is basic for the correct conception of ethics. An ethical theory that is concerned with well and ill-being is not an ethical theory in the proper sense, but a mere theory of prudence telling us how to act in order to have a pleasant life. An ethical theory in the proper sense tells us,

by contrast, what the difference is between good and evil actions.[17]

A second example is Frege's distinction between the concepts of being true and being taken to be true in his critique of the idealistic (or "psychologistic") foundation of logic.[18] These concepts, which are identified by the idealist logicians of his time, are independent according to Frege. What is taken to be true need not therefore be true. And it is not a contradiction that something which is taken to be true is actually false, he argues. In order to clarify the nature and the proper task of logic, it is essential, in his view, to distinguish clearly between these concepts. Thus, to discover the laws of truth is a task of logic, while to discover the laws of being taken to be true is not a task of logic, but of psychology.

The classical method to make the content of a given concept explicit is the decomposition of a concept into its conceptual constituents. Frege compares this method with the chemist's procedure to decompose a given substance into its molecular and atomic constituents. Just as the chemist analyses the chemical structure of substances, so too the philosopher analyses the logical structure of concepts. To display the logical structure of a given concept by means of an explication, the explicans must be syntactically more complex than the explicandum. Normally, the explicandum is syntactically simple and the explicans is composed of several predicates that are connected by logical signs. The meanings of the predicates occurring in the explicans are the conceptual constituents of which the explicandum is composed, and the manner in which the predicates are connected by logical signs represents the logical structure of the explicandum.

We may call explications whose method consists in the decomposition of concepts "analytic" explications. Obviously, this type of explication is applicable only to complex concepts. In order to make the content of an atomic (simple) concept explicit, we need an alternative method such as, for instance, the "holistic" method to determine the place of a given concept in our conceptual scheme. In this case the concept is not decomposed, but its conceptual connections with other concepts are made explicit. This procedure may be circular, i.e., it allows for the mutual explication of concepts. Thus,

[17] See Kant, 1788, p. 70.

[18] See the preface of Frege 1893.

Frege explicates the concept of judgment in terms of the concept of truth and vice versa: to judge is to acknowledge the truth of a given thought, and truth is the standard that we use to evaluate judgments in science.

If Donald Davidson is right, the attempt to decompose philosophical concepts into more fundamental concepts is a misguided project, because these concepts already belong to the most fundamental concepts we have. But this does not mean that these concepts cannot be explicated at all. In order, for instance, to explicate the concept of truth, which is considered by him to be primitive, we can adopt the holistic method to make its connections with other concepts like belief and linguistic meaning explicit.[19]

Quine also takes a critical stance on the conceptual program of traditional philosophy, but for another reason. In his view, it is impossible to distinguish objectively between "meaning and theory", between "analytic" and "synthetic sentences" and hence between conceptual analyses and empirical hypotheses.[20] Consider, for instance, the "folk theory of truth", which consists of the beliefs about truth shared by all competent speakers. This theory probably includes the beliefs that a sentence is either true or false, and that true sentences correspond to the facts. The problem posed by Quine's critique is that we do not have any method at our disposal that allows us to decide whether such a belief is analytic or synthetic. Since a conceptual analysis consists of analytic sentences, we consequently do not know how to decide whether an explication like "A true sentence corresponds to the facts" is to be considered as a conceptual explication or not. Applying the Moorean test, we might argue that this sentence is synthetic, because it is informative, in contrast to the tautological sentence "A true sentence is true". But this argument is not really conclusive because analytic sentences may also be informative. This is the case, for instance, when the components or the logical structure of the concept to be analysed are not obvious. Any explication revealing "hidden meanings" or "hidden semantic structure" is informative in some sense.

[19] Cf. Davidson 1996.

[20] Cf. Quine 1951 and Quine 1960, §§ 39 and 53.

4.2 Therapeutic explications

The second species of descriptive explications consists of the "therapeutic" explications, as we might call them alluding to Wittgenstein. These explications are also designed to make the content of concepts clear and distinct. However, in contrast to propaedeutic explications, their purpose is not to prepare the construction of a scientific discipline, but to show, on the contrary, that the endeavour to construct such a discipline is based on a linguistic misunderstanding.

An example of this type of explication is the minimalist account of truth advocated by Paul Horwich.[21] It falls into two parts. The first consists of an explication of the content and function of the word "true" according to which truth is a thoroughly formal, trivial and hence philosophically uninteresting concept that does not have any explanatory function in theories, but serves only certain technical needs that could also be achieved by quantifying over sentential variables. From this explication the consequence is drawn, in the second part of the account, that we do not really need a philosophical theory of truth.

Further examples of therapeutic explications are the non-cognitivist accounts of morality. Their purpose is to show that the endeavour to construct a moral theory is based on the linguistic misunderstanding that the predicate "is morally right" and its cognates are used to ascribe a property.

5 Revisionary explications

The context in which revisionary explications are made is the revision of theories. Their purpose is not to make the content of given concepts explicit, but to transform defective concepts into scientifically respectable ones. The defective concepts cannot simply be ignored because the play an important explanatory or expressive role in science. But they cannot be introduced directly into science because they are somehow defective. The aim of a revisionary explication is to revise the concept in such a way that, firstly, its explanatory or expressive function is preserved, and secondly, its defects are

[21] Cf. Horwich 1990.

eliminated.

Tarski's definition of truth is an example of this type of explication. According to him, the concept of truth is, on the one hand, indispensible for the construction of semantics and kindred disciplines, but, on the other hand, its direct introduction into science is illegitimate because of its inconsistency. The main purpose of his definition of truth is to transform the concept of truth into a concept that is both consistent and fruitful, that is, can be used to define the main concepts of logic and semantics like, for instance, the concept of logical consequence.[22]

The anti-realist explications of truth such as Dummett's verification theory are also of the revisionary type. They proceed from the assumption that the ordinary, realist concept of truth is scientifically not acceptable because it is potentially "recognition transcendent". If the truth of a sentence consists in its correspondence with reality, then there might be true sentences whose truth cannot be recognized, as, for instance, some undecidable sentences of set theory. To overcome this defect, the ordinary concept of truth must be replaced by a similar but epistemologically more adequate concept that can equally well "do the work" of the original concept, that is, that can be used to explicate the main concepts of logic and semantics (like logical consequence and linguistic meaning). The aim of the anti-realist explications of truth is to construct such a concept.

This example already shows that the realist- and anti-realist conceptions of truth are not necessarily incompatible. We may, for instance, accept the correspondence theory as a descriptive explication of truth, and the verificationist theory as a revisionary one.

In contrast to descriptive explications, a revisionary explication need not preserve the original meaning of the explicandum. To be materially adequate, it suffices that the meanings of the explicandum and the explanans are similar. To make this more precise, we need to explain what the conditions are for the explicandum and the explicans to be sufficiently similar with regard to their meanings.

To answer this question, some authors have proposed criteria re-

[22] The classical example of a revisionary explication is Kuratowski's definition of the concept of ordered pair in terms of the concept of set, which is considered by Quine as a "philosophical paradigm" (cf. Quine 1960, § 53).

ferring to the extension of the explicandum and the explanans. Carnap, for instance, adopted in the *Aufbau* the criterion of extensional identity. Goodman proposed the more liberal criterion of extensional homomorphism, and Quine the even more liberal criterion of extensional homomorphism.[23] These proposals ignore, however, what is essential for revisionary explications, namely, the preservation of the explanatory or expressive function of the explicandum. Hence, a revisionary explication should be considered as materially adequate if and only if the explicans can be used to perform this function.

Suppose, to illustrate this point, that the explicandum, say, the concept of truth, serves exclusively to refer to the members of its extension (true sentences) in a given theory. In this case the explicans must have, in order to fulfil this function, the same extension as the explicandum. If, on the other hand, the explicans is used also to refer to the concept of truth, then the explicans and the explicandum must have the same meaning.

The concept of number is a special case; its function is to count things. To fulfil this function, it suffices that the explicandum and the explicans are extensionally isomorphic. The reason is that every progression can be used to count things.

There are, however, cases in which even this liberal criterion is too narrow. The extension of the ordinary concept of existence is the set of existing objects. When this concept is explicated, as usual, as a concept of second order, the extension of the explicandum and the extension of the explicans do not even have any common member. The explicans does not apply to any object of which the explicandum is typically predicated. Nevertheless, this explication is materially adequate, at least when it is considered as a revisionary one. The defect of the explicandum is in this case that the truth-conditions of negative existence-statements like "Pegasus does not exist" are paradoxical, when existence is considered to be a concept of first order.[24]

These examples show clearly that the criteria of material adequacy for revisionary explications must be construed in a more flexible and differentiated way. Our criterion that a revisionary explication

[23] Cf.Carnap 1928, §§ 43-45, Goodman 1978, pp. 99-102, Quine 1964 and Quine 1969, pp. 55.

[24] Cf. Quine 1948.

must preserve the function of the explicandum takes this into account. Examples of explications that are not materially adequate on this criterion are the traditional explications of truth from the point of view of the disquotation theory. For, these explications are based on the assumption that the truth-predicate serves to ascribe a property. But, according to the disquotation theory, the truth-predicate cannot be used to this end, because it has a very special function, namely, to disquote quoted sentences.

A second shortcoming of the current debate about the adequacy of revisionary explications is that the criteria of adequacy discussed are incomplete. Thus, the adequacy of such explications depends also on the adequacy of the norms of scientific respectability that the explicans is supposed to satisfy. Some examples may illustrate this point.

The supreme norm in Frege's *Grundgesetze* for the methodologically sound introduction of new expressions into science proper is this: we have to "explain each expression with respect to its reference completely" (Frege 1903, §§ 57, 65). According to this norm, the use of an expression in science proper is legitimate only when its referential indeterminacy has been removed. This restriction is motivated by Frege's quest of ensuring the validity of the logical laws in rigorous science, in particular the validity of *tertium non datur*.[25]

A second norm due to Frege is the demand for a proper criterion of identity for abstract singular terms. "If we are to use the symbol *a* to signify an object", he writes, "we must have a criterion for deciding in all cases whether *b* is the same as *a*" (1884, § 62). This norm, which was later adopted by Quine, has come to be known as the norm "No Entity without Identity".[26]

Obviously, a revisionary explication that aims to adapt the explicandum to these norms is adequate only insofar as these norms are adequate. With regard to Frege's norm of referential determinacy, this is highly questionable, however, because it is impossible to determine the reference of all terms completely.

Secondly, the adequacy of a revisionary explication depends also on the adequacy of the method of explication it employs. This "meth-

[25] Cf. Frege 1903, §§ 55 ff.

[26] For a reconstruction of this norm, see Greimann 2000.

odological adequacy" of an explication depends, in the first place, on the norms of scientific respectability that are adopted. Suppose, for instance, that we want to transform the inconsistent ordinary concept of truth into a consistent concept. To ensure the consistency of the explicans, we must replace the explicandum by concepts whose consistency is beyond question. To this end, we may, for instance, follow Tarski's strategy to reduce the concept of truth to physical concepts.[27] Consequently, an explication seeking to transform truth into a consistent concept is methodologically adequate only if the method employed achieves a reduction (or "elimination") of the explicandum. The explicit definition of the explicandum is an example of a method satisfying this constraint, and the axiomatic characterization of the explicandum is an example of a method that does not satisfy it.

Thirdly, the adequacy of a revisionary explications depends also on the adequacy of the order of explication on which it is based. Such an order tells us which kind of concepts is to be explicated in terms of which other kind of concepts. Its task is to ensure that the scientifically defective concepts are explicated in terms of more respectable concepts, and not vice versa. Frege's explication of the concept of number in terms of logical concepts, for instance, is based on an order of explication according to which logical concepts are scientifically more respectable than arithmetical ones. To explicate logical concepts in terms of arithmetical ones would be a case of *clarum per obscurum*, from his point of view.

Finally, the adequacy of a revisionary explication depends also on its form, its structure. To be formally adequate, the explication must include all measures that are necessary to explicate the explicandum in a satisfactory way. Moreover, these measures must be taken in the correct order. The first step must be to determine the explicandum of the explication. Secondly, it must be shown that the explicandum is defective. Thirdly, the function to be preserved must be made explicit. Fourthly, a list of norms of scientific respectability must be set up and justified. Fifthly, an order of explication must be established that fixes the range of the possible explanantia. Sixthly, a method of explication must be selected by means of which the explicandum

[27] For more details, see Etchemendy 1988.

can be adapted to the norms of scientific respectability. Seventh, this method must be applied to the explicandum. Eighthly, it must be shown that the explicans satisfies the norms of scientific respectability.[28] As far as I can see, there is only one explication to be found in the literature that is formally correct in this extensive sense, namely, Tarski's definition of truth, which can therefore be considered as a paradigm of a revisionary explication.

6 Reductive explications

There are many programs in philosophy who are not concerned with the explication of a single concept, but with the explication of whole families of concepts. Examples of this type of explication are the reductive programs in the philosophy of mind, the philosophy of science and the philosophy of mathematics, like (analytic) behaviourism, phenomenalism, physicalism, and logicism. These programs do not aim to describe the structure of a single concept, but the structure of our overall conceptual scheme. According to (analytic) physicalism, for instance, our conceptual scheme has the hierarchical structure of a building whose fundament consists of the physical concepts.[29] All other families of concepts such as the psychological, sociological and semantic concepts occur at higher levels, and they can be reduced to the fundamental concepts via chains of explicit definitions.

The main purpose of a reductive explication of a given concept is neither to make its content explicit nor to transform it into a more respectable concept, but to show that it is in principle *superfluous*, in the sense that it can always be substituted by the explicans. An explication of this type may be considered as a rule for the translation of a formulation of a given theory that contains the explicandum into a formulation of the same theory that contains the explicans, but not the explicandum. Such a rule shows how the explicandum can be "eliminated", by showing how every sentence containing the

[28] In Greimann 2007, the formal adequacy of explications is described in more detail.

[29] There are weaker versions of physicalism that do not involve this thesis. They are not relevant in the current context.

explicandum can be translated into a sentence that does not contain it. In the case of logicism, for instance, the source language is the arithmetical language and the target language the language of logic.

To show that the explicandum is superfluous, it suffices to show that the function it performs in the theory can be performed equally well by the explicans. The criterion of material adequacy for reductive explications is therefore identical with criterion for revisionary explication: the function of the explicandum must be preserved.

The redundancy theory of truth can be considered as a reductive explication of the concept of truth. As already indicated by its name, the core of this theory is the doctrine that the word "true" is superfluous. The rule of translation it suggests for the translation of sentences in which "true" occurs is this: a sentence of the form *It is true that p* or *"p" is true* is to be translated simply as *p*.[30]

Tarski's definition of satisfaction and truth in terms of physical concepts (including set-theoretical ones) is a second example of a reductive explication. Unfortunately, this explication is not materially adequate because it does not preserve the intended function of the explicandum. For, the explicans is supposed to play an important explanatory function in the foundation of semantics, viz. the definition of the concept of logical consequence and kindred notions. This presupposes that the explicans can be used to explain the semantic truth-conditions of sentences. Actually, however, it cannot be used to this end, because the explication implies that a semantic explanation like "The sentence "Snow is white" is true if and only if snow is white" says the same as the tautological explanation "Snow is white if and only if snow is white".[31]

A second criterion of adequacy for reductive explications is that the translation rule must be general. In particular, the translation rule should be applicable to "new cases". To see this, suppose that the source language consists of exactly the following four sentences:

[30] To translate the quantified occurrences of "true", as in "All beliefs of Plato are true", we need to quantify over sentential variables. Thus, we can translate "All beliefs of Plato are true" as "For all p: if Plato believes that p, then p".

[31] For a more thorough discussion of this problem, see Etchemendy 1988.

Snow is white,
Snow is black,
"Snow is white" is true,
"Snow is black" is true".

In this case, we may translate the sentences containing "true" into a target language that does not contain this word by means of the following list-like explication:

> x is true if and only if x = "Snow is white" and snow is white, or x = "Snow is black" and snow is black.

Though this explication does enable us to eliminate the truth-predicate, it is unsatisfactory because it does not show us how to eliminate the truth-predicate when it is applied to new sentences that we may introduce into the source language.[32] To overcome this difficulty, we need a more general rule of translation that accounts for such new cases.

7 Stipulative explications

Roughly speaking, a descriptive explication informs us about what the explicandum means in natural language, and a revisionary explication tells us what the explicandum should mean in science. The characteristic of a stipulative explication is that it informs us about how the author or speaker wishes the explicandum to be understood when she or he uses it.

Suppose, for instance, that an author wants to explicate the classical, Aristotelian concept of truth. In order to make clear that his explication refers to this concept of truth and not to another, he may stipulate that by "truth" he means "truth in the classical, Aristotelian sense". This stipulation is an example of a stipulative explication. A second example is Tarski's proposal to denote the classical, Aristotelian concept of truth by the artificial term "frue", if the majority of the theoreticians of truth should decide that the word "true" should be reserved to denote a different concept of truth.

Generally speaking, the purpose of a stipulative explication is to

[32] This criterion is suggested by Hartry Field's criticism of Tarski's definition of truth in Field 1972.

fix the intended meaning of the explicandum in order to use it as a technical term. The explications of this type divide into two species.

7.1 Explications of laziness

The main purpose of an explication of laziness is to abbreviate a longer expression. Thus, the stipulation to understand by "truth" the classical, Aristotelian concept of truth allows us to abbreviate the longer expression "truth in the classical, Aristotelian sense" by "truth".

From a theoretical point of view, explications of laziness are superfluous. They are relevant only for the linguistic quality of a theory formulation. This quality depends, for instance, on the readability and the elegance of the formulation. The criteria of adequacy for explications of laziness refer correspondingly to pragmatic and aesthetic qualities such as brevity and elegance of expression.

7.2 Metaphorical explications

The second species of stipulative explications consists of the "metaphorical" ones. Their purpose is to attach a new meaning to the explicandum that cannot be expressed by means of the expressions that are available in the language. In contrast to explications of laziness, metaphorical explications extend the expressive power of the language in an essential way.

An example of this type is Frege's explication of the concept of a "truth-value". According to it, the truth-value of a sentence is the circumstance of its truth or its falsity.[33] Its purpose is to inform the reader about what Frege wishes to understand by the "truth-value" of a given sentence or thought.

Syntactically, "the truth-value of x" is a functional expression. Applied to a proper name, it forms a more complex proper name. Semantically, it attributes to a given argument its truth-value, just as "the colour of x" attributes to a given argument its colour and "the temperature of x" attributes to a given argument its temperature. In natural language, we can express that the colour of snow is

[33] Cf. Frege 1892.

whiteness, but we cannot express that the truth-value of "Snow is white" is truth. The problem is that there is no expression of natural language that expresses the intended meaning of "truth-value". The purpose of a metaphorical explication is to fill such gaps. It does not abbreviate a longer expression, but provides a new meaning for an old expression.

Since the new meanings cannot be expressed by means of the expressions that are already at our disposal, it is strictly speaking impossible to fix the intended meaning of the explicandum. The only resources we have to explain what a truth-value is supposed to be are the expressions of natural language.[34]

For this reason, the success of a metaphorical explication depends on the imagination and the empathy of the hearer or reader. He or she must in a certain sense *guess* the intended meaning of the explicandum. The method we commonly use to overcome this difficulty is the metaphorical use of natural language terms. Thus, the designation of the truth-values as "truth-values" is supposed to give hints to the reader that enable him to guess or to infer in some sense the intended meaning of the explicandum. This metaphorical method is also used in the formation of concepts like "logical decomposition", "concept of second order", "analytic explication", "ontological commitment", "hypothetical imperative", and so on.

Since the majority of the technical terms used in philosophy are introduced in this way, the metaphorical explication is perhaps the most important type of explication in philosophy. Only very few concepts are formed by Quine's method of elimination (or reduction).[35] For this reason, the current debate about explication should pay more attention to this type.

Although metaphorical explications are based on stipulations, they are not arbitrary. In contrast to explications of laziness, a metaphorical explication can be successful only when the metaphorical meaning of the explicandum can somehow be derived from its literal meaning. Consequently, an explication of this type is materially ade-

[34] The application of ostensive definition and kindred methods is impossible in this case, because truth-values are abstract objects.

[35] In Greimann 2009, I have tried to show that Quine's ontological reductions by means of contextual definitions do not really work.

quate only when the intended meaning of the explicandum is similar or analogous to its literal meaning. Otherwise the explication would not contain any hints that the hearer or speaker needs in order to guess what the intended meaning is supposed to be.

Since the success of a metaphorical explication depends strongly on the imagination, the empathy and the scientific socialization of the hearer or reader, its material adequacy must be relativized to linguistic communities. An explication that is materially adequate for one community, say, the readers of Frege, may be inadequate for another community, say, the readers of Heidegger.

8 The adequacy of the suggested typology

I have tried to determine the main types of conceptual explications. The main criteria of classification I used to this end refer to the objects and the purposes of explications. With regard to the objects, we must distinguish between nominal and real explications, and between intensional and extensional ones. With regard to their purposes, conceptual explications divide into five main groups: descriptive, revisionary, reductive and stipulative definitions. The descriptive explications can be subdivided into propaedeutic and therapeutic ones, and the stipulative explications into explications of laziness and metaphorical explications. In order to refine this typology, we could classify explications with regard to their methods and their form.

The adequacy of the suggested typology is not a question of its truth or falsity, but of its utility. It may be useful mainly with regard to the following three ends.

First, it enables us to evaluate the adequacy of a given explication in a more sophisticated way. Consider, for instance, Tarski's definition of truth. It does not consist of a single explication, but of various different explications that serve different purposes and have different criteria of adequacy. The first step of Tarski's explication consists of a stipulative explication by means of which the intended meaning of the definiendum is fixed. With regard to its purpose, the definition involves a descriptive explication of truth that is supposed to "capture" this meaning, the classical concept of truth. At the same time, the definition contains a revisionary explication that is supposed to replace the ordinary concept of truth by more respectable concepts.

With regard to its object, the definition must be considered as an extensional explication whose object is the set of true sentences of the object language. Nevertheless, it can also be considered as an intensional explication designed to explain the meaning of "true" from the classical point of view. Finally, the definition can also be considered as a real definition telling us on which language-world relations the property of truth depends. In order to evaluate Tarski's definition adequately, we must take into account that it contains various explications of very different types and with different criteria of adequacy. It might turn out that some of these explications are satisfactory and some not.

Second, our typology may contribute to the task tackled by Kirkham to resolve some of the confusions to be found in the current debate about truth. The typology shows, for instance, that the realist and the anti-realist theories of truth must not necessarily be seen as competitors, because they may also be regarded as complementary explications with different goals and different criteria of adequacy.

Finally, and most importantly, the typology may also contribute to the project envisaged by Carnap to construct an explicit theory of conceptual explications in which the rules for the correct explication of concepts are discussed.[36] To discuss these rules, we need to know the criteria of adequacy for explications. And to explain these criteria in an adequate way, we must distinguish between different types of conceptual explications, because the criteria of adequacy of a conceptual explication depend on its type.[37]

Dirk Greimann
Fluminense Federal University

References

Alston, William. 2002. Truth: Concept and Property. In *What is Truth?*, ed. by Richard Schantz. Berlin, New York: de Gruyter, pp. 11-26.

[36] In Greimann 2007, a system of rules for the correct revisionary explication of concepts is developed.

[37] This paper is based on a talk that I delivered at the Universität Hamburg, the Universität Münster and the Universidade de São Paulo in 2010. I am indepted to Guido Imaguire for his suggestion to write a paper on this topic.

Brandom, R. 1998. *Making it Explicit*. Second edition. Cambridge/M.: Harvard University Press.
Brink, David. 1989. *Moral Realism and the Foundations of Ethics*. Cambridge/M.: Cambridge University Press.
Carnap, Rudolf. 1928. *Der logische Aufbau der Welt*. Berlin.
Carnap, Rudolf. 1950. *Logical Foundations of Probability*. Chicago.
Davidson, Donald. 1996. The Folly of Trying to Define Truth. *Journal of Philosophy* 93, 263–278.
Etchemendy, John. 1988. Tarski on Truth and Logical Consequence. *Journal of Symbolic Logic* 53, 51–79.
Field, Hartry. 1972. Tarski's Theory of Truth. *The Journal of Philosophy* 69, pp. 347-375.
Frege, Gottlob. 1884. *Die Grundlagen der Arithmetik. Eine logisch-mathematische Untersuchung über den Begriff der Zahl*. Breslau, reprint: Darmstadt, 1961.
Frege, Gottlob. 1892. Über Sinn und Bedeutung. In *Funktion, Begriff, Bedeutung*. Göttingen: Vandenhoeck und Ruprecht, 1980, pp. 40-65.
Frege, Gottlob. 1893. *Grundgesetze der Arithmetik. Begriffsschriftlich abgeleitet*, Vol. I. Jena, reprint: Darmstadt, 1962.
Frege, Gottlob. 1903. *Grundgesetze der Arithmetik. Begriffsschriftlich abgeleitet*, Vol. II, Jena, reprint: Darmstadt, 1962.
Frege, Gottlob. 1914. Logik in der Mathematik. In *Nachgelassene Schriften*, ed. by H. Hermes, F. Kambartel and F. Kaulbach. Hamburg: Felix Meiner, 1983, pp. 219-270.
Goodman, Nelson. 1978. *Ways of Worldmaking*. Indianapolis: Indiana.
Greimann, Dirk. 2009. Contextual Definition and Ontological Commitment. *Australasian Journal of Philosophy* 87, pp. 1-17.
Greimann, Dirk. 2007. Regeln für das korrekte Explizieren von Begriffen. *Zeitschrift für philosophische Forschung* 61, pp. 261-282.
Greimann, Dirk. 2000. No Entity without Identity: A Reductionist Dogma? *Grazer Philosophische Studien* 60, pp. 13–29.
Gupta, Anil. 2008. Definitions. *Stanford Encyclopedia of Philosophy*, online edition, http://plato.stanford.edu/entries/definitions.
Horwich, Paul. 1990. *Truth*. Oxford: Basil Blackwell.
Kant, Immanuel. 1788. *Kritik der praktischen Vernunft*. Hamburg: Meiner, 1990.
Kirkham, Richard. 1995. *Theories of Truth. A Critical Introduction*. Cambridge/M.: Cambridge University Press.
Quine, Willard van Orman. 1948. On What There Is. In *From a Logical Point of View*. Second edition. Cambridge, Mass.: Harvard University Press 1964, pp. 1–19.
Quine, Willard van Orman. 1951. Two Dogmas of Empiricism. In *From a Logical Point of View*. Second edition. Cambridge/Mass. Harvard University Press, 1964, pp. 20-24.
Quine, Willard van Orman. 1960. *Word and Object*. Cambridge/M.
Quine, Willard van Orman. 1964. Ontological Reductions and the World of Numbers. In *The Ways of Paradox and Other Essays*. Second edition. Cambridge/M., pp. 212–220.
Quine, Willard van Orman. 1969. Ontological Relativity. In *Ontological Relativity and Other Essays*. New York, pp. 26–68.
Tarski, Alfred. 1935. Der Wahrheitsbegriff in den formalisierten Sprachen. In *Logik-Texte*, ed. by K. Berka and L. Kreiser. Darmstadt, 1983, pp. 445–546.
Tarski, Alfred. 1944. The Semantic Conception of Truth and the Foundations of Semantics. *Philosophy and Phenomenological Research* 4, pp. 341-376.

On Identifying and Identification

Breno Hax, Jr.
Federal University of Paraná

BIBLID [0873-626X (2012) 34; pp. 671-687]

The first aim of this paper is to sketch an account both of the process of identifying and of identifications, the results of the process. I treat identifications as sorts of files of information, or *identification files*. The role of identification files as devices for use in reidentification is central to my account. The process of identifying x is explained in terms of the perception of x and the formation of an identification file that has x as its topic. The topic is the entity the identification file is about.

My proposal does not aspire to capture every aspect of the usage of 'identification' in ordinary philosophical discourse. 'Identification' is a theoretical term, as noted by G. Evans[1]. My proposal should be judged by its theoretical consequences.

The second aim of this paper is to outline the way how the cognitive system (or mind) yields two remarkable kinds of correlation between identifications and topics: equivocal identifications (identifications with two or more topics) and, in addition, different identifications of the same topic. To illuminate the issue, I introduce the distinction between *vertical* and *horizontal* information connecting. I propose the notion of *supermap* as a useful metaphor for explaining the nature of information connecting by mind.

Let me make a concise observation about thought and reference. I am providing an account of identification that aims to be serviceable for theories of thought and reference. It is plausible that a full theory of how a thought is about something rather than anything else needs some account of identification. What makes a thought be about a cer-

[1] Evans (1982: 145).

tain entity x? This seems to be a job for the notion of identification. It seems prima facie plausible that having an identification of an item x is a necessary condition for an agent A to be able to think about x. In addition, if in order to refer to an entity we need to think about it, theories of reference need to include identification in some way in their accounts. The arguments advanced here will not depend on such assumptions about thought and reference.

This paper will be deliberately silent on whether identifications are concepts or devices working associated to concepts (sometimes called 'conceptions').

The structure of the paper is as follows. Section 1 introduces my proposal about identifying and identification. My account treats identification in terms of perception and information. Section 2 examines the notion of information. Section 3 deals with perception. Sections 4-6 discuss the nature of identification and its connection to reidentification. Section 7 presents two facts which any adequate theory of identification must respect. Finally, section 8 presents the distinction between vertical and horizontal information connecting. The notion of supermap is introduced.

1 Identification: process and result

'Identification', like many terms, is ambiguous between meaning a process and meaning the product of that process. My proposal concerns the two meanings of 'identification': the *process of identifying* and the results of the process, the *identifications*.

The results from the process of identifying can be treated as files. I shall call them 'identification files'. The *topic* is the entity (in the most general sense of the word 'entity') which the identification file is about[2].

[2] The term 'topic' is used in the same sense by Woodfield 1991. The nature of the relation between identification file and topic is not the theme of this paper. I limit myself to point out that there are different particular relations linking identifications and topics. Identification files triggered by perception present their topics due to an informational relation – the received information is information about the topic. Identification files formed by imagination present their topics due to some kind of fit between the conditions included in the identification and the topic. It is plausible that the relation between identification files created by

What *are* identification files? How may the metaphorical notion of a file help us in understanding the nature of identifications? Let me use the notion of information to sketch an answer. Human beings, like other kinds of organisms, capture information about the world through perception. This information is inserted into and stored in our cognitive system. Identification files are cognitive unities constituted by pieces of information that the cognitive system treats as devices for tracking something in reality. I develop this idea in a greater depth in section 4.

I shall classify a process as an identifying process if it begins with the perception of a certain entity and leads in an appropriate way to the formation of an identification file of the same entity. In my proposal, an agent A identifies an entity x iff:

1. A perceives x;
2. The process of A perceiving x results in an appropriate way in the formation of an identification file of x.

The condition (1) excludes from the extension of 'identifying' any process of formation of an identification file that does not begin with the perception of the topic of the identification. The formation of the identification file must be triggered by the perception of the entity which is its topic. The relation between the perceived entity that triggers the identification file and its topic is the relation of identity.

The process of identifying is not the only way of producing identification files. I shall call identification files acquired by the identifying process *perceptual identification files*. Besides them, we have identification files of entities not perceived and so not identified. We may imagine an object by positing something that combines properties – for example, the man with the property of being the fastest cyclist in the world and the property of being the best mathematician of the world. As a result, we have formed an identification file but it does not seem correct to say that we identify its topic. It is a case in which we form the identification file without identifying its topic. The same happens in communication episodes. We can acquire iden-

---

suspicion and topics is informational as happens with perceptual identifications. Cf. section 5 about identifications created by suspicion.

tification of some entity only by hearsay. However, the hearer does not identify something merely by hearing about it[3]. Section 5 will expand on the theme of formation of identification files without any identifying process.

My account treats 'to identify' as a success verb. I propose that it is correct to say that we identify x uniquely if x exists and we perceive x. The verb 'to identify' should be treated as a success verb because of the connection between the identification act and the reidentification act. After all, the verb 'to reidentify' is a success verb. It seems appropriate to speak in terms of reidentifying only in cases when we meet the entity again. This involves the satisfaction of two external conditions: (i) the entity can exist independently of our meetings and (ii) be met in different episodes. Our alternatives to treating 'to identify' as a success verb would be severing the connection between identification and reidentification (which I claim is a non-starter) or depriving the reidentification notion of success conditions. The last alternative would trivialize the notion of reidentification. The notion of reidentification has an important role in a description of the interaction between agent and world because of its success conditions.

Given the connection between identification and reidentification, only the treatment of 'to identify' as a success verb allows us to ensure the non-triviality of the use of the notion of reidentification. If to identify were merely a cognitive act without conditions of success, what point would there be in talking about reidentifying something? If, for example, merely thinking about Pegasus were an act of identifying Pegasus, then when were I in the future to think again about Pegasus, would I be reidentifying Pegasus?

I propose treating the notion of meeting something in terms of perceptual meeting. The inclusion of the condition of perception in my account restricts the conditions for something to be identified:

[3] In a theory of perception that claims that the formation of an identification file is a condition for a successful process of perception, the condition (2) would be involved in the condition (1). According to such a theory, the formation of an identification would not be triggered by perception but it would be part of the process called 'perception'. I believe that the problem of such a theory is that it incorrectly fuses perception and registration of information acquired by perception.

besides existing, the entity shall be perceived[4].

The notions of perception and identification file can be defined in terms of the notion of information. To perceive is to gather information about the environment. Identification files are clusters of pieces of information that the cognitive system treats as cognitive modules about something. I shall briefly discuss the notions of information and perception and then return to discussing the nature of identification files.

In what follows, I shall use the term 'identification' as short for 'identification file'.

2 Information

I believe that the treatment of the notion of information provided by Fred Dretske is correct in its essential features[5]. This treatment is nicely captured in John Perry's formulation: 'Information is basically what one part or aspect of the universe (the signal) shows about some other part or aspect (the subject matter)' (Perry 2002: 175). One aspect of the universe is capable of carrying information about another because both are related by some constraint, law or principle. If the aspects of the world were not interrelated by constraints, nothing would be able to carry information. I take constraints as necessary or probabilistic relations among types of states of affairs.

A signal contains information about what has to be the case for it to happen. If there is a non accidental connection between being F and being G such that, for example, being F is a necessary condition for something being G, then the fact that x is G carries information that x is F. If it were merely coincidence that what is G is F too, then the fact of x being G would not carry information that x is F.

For example, every grade of expansion of the volume of a portion of mercury is a type of state of affairs that depends on another type of states of affairs, the degree of temperature of the mercury (and its

[4] One might object that my account is excessively restrictive because the meetings could be non perceptual ones. Any meeting with an entity in which the agent receives information about it and forms an identification file for detecting it in future meetings will count as perceptual in my sense. Cf. sections 2 and 3.

[5] Cf. Dretske 1981.

pressure). If the temperature were not raised, the mercury's volume would not expand (at constant pressure). Due to this connection, the degrees of expansion of the mercury's volume embody information about its temperature.

A state of affairs carries information about another state of affairs only if it is not merely a case of correlation or co-occurrence but involves dependence. A certain state of affairs carries information about the state of affairs on which it depends.

Besides nomological or logical dependence, three other features of information are essential to my treatment of identification:

(a) Agent-independence: information is independent of agents that collect it. The world is full of information because information depends only on the existence of constraints between types of states of affairs.
(b) Portability: information can be acquired, stored and transmitted. This makes the flow of information possible. Additionally, the process of capturing information does not change it.
(c) Aboutness: information is about its source independently of being captured by our senses. For example, the pattern of the rays of the light reflected from the situation that *a is F* carries the information that *a is F*[6]. As reflection patterns depend nomically on surfaces, this pattern carries information about that situation[7].

3 Perception and uniformities

I shall take three things for granted.

First, there are individuals in the world having properties, standing in relations and occupying spatiotemporal locations.

Second, individuals, properties, relations and locations can occur in different combinations. In this manner, individuals, properties,

[6] The example is from Adams 1999.

[7] Cf. the Dretske's discussion of intentionality as a feature of the physical world in Dretske (2000: 209-12).

relations and locations can be treated as *uniformities*[8]. Every possible combination is a state of affairs. *Situations* are the actual states of affairs.

Individuals are uniformities recurring in different situations because they persist at least for some time holding properties and relations. Properties and relations are uniformities because they are exemplified by different individuals in different locations. Locations are uniformities because they can be occupied by different entities. A place can be occupied by different individuals at different times. Different individuals can be in different places at the same time.

Third, human beings are able to perceive some uniformities of the environment. It is plausible that different kinds of biological organisms are able to perceive different uniformities. Surely, the uniformities perceived by every kind of organism are in accordance with their members' need to survive. Likewise, it is highly plausible that different kinds of organisms are able to detect a large amount of the same uniformities.

I outline in this paper a proposal on identification capable of coping with the three points above by appropriately connecting identification and perception.

I think that a correct account of identification needs to capture the following fact: the entities that we identify and reidentify are the uniformities that we perceive. Thus, in my view, a correct account of identification needs to be supported by a theory of perception capable of selecting environmental uniformities as perceptual objects. Informational theories of perception like Dretske's are what we need here[9].

Note that someone can accept my proposal that the topics of identification are the perceptual objects and, even so, not consider that informational theories are capable of correctly determining the perceptual objects. I shall not argue in favor of informational theories but I shall just point out that their outcomes are taylor-made for a theory of identification.

One fundamental feature of Dretske's theory of perception is that

[8] I am using the notion of uniformity of Barwise and Perry (1999: 8).

[9] Dretske (1981: chapter 6). Dretske talks about invariants. For the present purposes, I take 'invariants' and 'uniformities' as applying to the same items.

it enables us to explain why we perceive human beings, trees and stones without perceiving the changing retinal images that are causal links in the delivery of information. For Dretske, perception carries information about, for example, the situation that *a is F* and not about the perceptual events which operate as information-carrying causal links because the situation that *a is F* has a higher order of invariance in contrast to variant perceptual events[10]. Thus, perception carries information about its constant source without carrying information about the more proximal and variable events of the causal chain through which the information is carried.

Another advantage of adopting an informational account of perception such as Dretske's is to acquire protection against the relativist consequences of conceiving identification as theory-laden or belief-laden. The class of views that subscribe to the thesis that identification is theory-laden is committed to the claim that if you exchange the theory adopted by the agent for another theory, then – *voilà*! – the agent identifies a different kind of entities. Thus, oddly enough, identification would be relative to the agent's theory. In accordance to Dretske's theory, the perceptual object is selected independently of the beliefs and theories of the perceiver. To perceive trees does not depend on our theories and beliefs. We perceive such entities because we receive information about them. The adoption of this view of perception in combination with the thesis that what is perceived is what is identified can block the claim that every act of identification is theory-laden or belief-laden[11].

4 Identification and uniformities

In section 1, I advanced the proposal that identifying x consists of perceiving x and forming an identification file which has x as its topic. Now it is time to characterize what an identification file is.

The metaphor of identification files is a logical one. The infor-

[10] Dretske (1981: 145).

[11] I really think that this sort of relativist theories relies upon theories of perception which adopt the thesis that the beliefs of the agent (or his theories) determine the perceptual object. For a refutation of this class of theories, cf. Dretske (2000: 105-7).

mation an identification file contains is not physically separated in a module. For example, my identification file of a friend of mine is composed of pieces of information that are physically registered in different parts of my brain. The cognitive system (or mind) treats them as forming a kind of cognitive unity. I will not try to account for how the cognitive system establishes this unity. I think that the mere phenomenon of this unity is a clear sign that the cognitive system is intentional. It works to unify pieces of information in terms of the item or items they are about. Thus, an identification is a cluster of pieces of information that the cognitive system treats as being about a certain entity.

The workings of the mind to unify pieces of information about the same entity are not always successful. One thing is the aboutness of a piece of information, the 'informational aboutness'. Another one is the aboutness ascribed by the mind, the 'conferred aboutness'. This last one determines which pieces of information are inserted in an identification. I take as basic the mind's capacity to treat something as being about something else.

Individuals, properties, relations and locations are recurring. The primary function of identifying is to insert information (received by the senses) about recurring uniformities (the location of food, the appearance of a dangerous kind of animal, etc.) in the cognitive system to use in future meetings with them.

The point of having identifications is to make it possible to track their topics. If there were no reidentification, the memory faculty would merely store pieces of information of no future use. These pieces of information would be taken as pieces of information about entities encountered only once.

Given that identification files are used to detect and track the uniformities of the world, we shall not conceive of them as sums of pieces of information, but as flexible devices of integrated information presenting the topic so as to enable the organism to detect it. In section 8, I discuss the way the cognitive system deploys networks of identification files to detect and reidentify items.

5 Identifications without identifying, ur-identifications, and derivative identifications

We also form identifications without the identifying process. My objective in this section is to sketch how my account accommodates this kind of identifications.

I follow Evans in proposing that perception, memory and communication compose the 'informational system'[12]. The information about x captured by identifying x flows across the community of speakers by communication. Communication allows the transmission and acquisition of identifications formed by another agent. If the identification transmission is successful, the receiver acquires an identification of the same topic.

One speaker can acquire his identification from another speaker in a communication process as a user of a language. Identifications can take two distinct positions in a communicative net. *Derivative identifications* are the identifications derived from identifications formed by another speaker. A non-derivative identification is an *ur-identification*.

The *producer* is the agent who forms the ur-identification of a certain topic. In the communication process, the producer transmits information of his identification to other speakers. *Consumers* are speakers who acquire the identification by communication[13].

Identification files acquired by communication appropriately linked to perceptual ur-identifications preserve the same link to the entity the information is about. That is, they have the same topic as the ur-identification. Transmissions of identifications in a communicative net are topic preserving.

Consumers' identifications are informationally dependent on producers' identifications. Typically, consumers use their identifications for representing the same topic of the ur-identification on which it depends.

Concerning the origin of identifications, a *perceptual identification*

[12] Evans (1982: 122).

[13] The distinction between producer and consumer is inspired by Evans's distinction between introducers of a name and users who acquire it as mere participants in the linguistic community. Cf. Evans (1982: chapter 11).

is one formed by the activity of identifying its topic. According to my account, perceiving the entity which the identification is about is a necessary condition for identifying it. Identifications of topics unperceived by anyone are *non perceptual identifications*[14]. Ur-identifications can be perceptual identifications or non perceptual identifications. A derivative identification is perceptual if and only if the ur-identification from which it originates is perceptual.

In my view, identifications formed in episodes of illusion are perceptual ones. After all, there is something perceived albeit with interference from distorting factors. The resulting identification is of the perceived item.

Non perceptual identifications can be brought about by imagination, suspicion or hallucination.

Imagination cases are cases of formation of an identification by a combination of characteristics of entities. The topic in this case, if any, is what fits the discriminating information. Identifications of fictional personages are examples of such a process.

The suspicion case is that in which the producer has clues, evidence or reasons for positing that something exists although unperceived.

For example, a man sees (what he takes as) footprints and comes to suspect that they are caused by a certain entity. This suspicion triggers the formation of an identification for tracking the entity, if any, that made the footprints. One real example was the postulation by Urbain Jean J. Leverrier of the existence of Neptune on the basis of the observed orbit of Uranus. Leverrier suspected, given his observations, that a celestial body was causing the perturbations in the orbit of Uranus. He formed an identification of the celestial body and associated it to the name 'Neptune'. In such cases, the ur-identification's topic is merely posited, rather than perceived.

In hallucination, there is no perceived item, but the agent assumes that there is one due to the abnormal working of the sensory and cognitive faculties. The resultant identification is non perceptual be-

[14] According to my proposal of including perception as a necessary condition for identifying, the class of identifications acquired by the identification process and the class of perceptual identifications are co-extensives. In accounts which do not include perception as a necessary ingredient for the identification process, the two classes can differ in extension.

cause there was no perception at all.

6 Reidentification

Whereas 'identification' has a process/product ambiguity, 'reidentification' designates only a process, the activity of reidentifying[15]. Such activity presupposes a successful occurrence of a previous activity of identifying.

The act of reidentification does not produce special types of identification files. It only sophisticates and increases the amount of information gathered in the identification files previously acquired. The formation of a new identification file would be a new act of identifying and not an act of *re*identifying. This avoids the usage of 'reidentification' to signify a product and thus eliminates any ambiguity between process and product.

There are two relevant conditions for the correct application of the verb 'to reidentify', one internal and another external.

The internal condition is that the agent shall previously have a perceptual ur-identification of x to be capable of reidentifying x. This avoids the possibility that the first perceptual meeting of an agent with x counts as a reidentification of x. If Marco Ruffino acquires an identification of Nathan Salmon exclusively by communication and by reading books about philosophy, it is not correct to say that he has identified Nathan Salmon. For the same reason, it is not correct to say that when Marco Ruffino subsequently is presented by a friend to Nathan Salmon himself, he reidentifies Nathan Salmon.

The external condition for there to be reidentification is a stringent ontological condition. There is only reidentification if there is numerical identity between the entity previously identified and the entity encountered now. In other words, the activity of reidentification can be successful only if we have different meetings with the same entity.

[15] My use of 'reidentification' is not equivalent to the Strawson's use in Strawson 1959.

7 Two facts

Any correct account of identification has to respect two general facts: what I refer to as the 'externalist fact' and the 'opacity fact'. In this section I present and discuss these facts. I suggest an explanation for the first fact. The sketch of a proposal to explain the second fact is made in the next section.

> *The externalist fact*. The perceptual identification is *of* the entity that is the source of its information[16].

I treat the externalist fact as a consequence of the nature of the information. The intentional nature of the information explains the connection between perceptual identification and its topic. The explanation can be presented as follows:

1. Every piece of information is about its source.
2. Identifications are composed of pieces of information.
3. Perceptual identifications inherit their aboutness from the aboutness of the pieces of information that compose them.
4. Perceptual identifications are identifications of the entity that is the source of its pieces of information.

Information that *a is F* is information about *a*. Consider the identification composed only of the information that *a is F* and the information that *a is G*. Both pieces of information are about *a*. As a result, the identification composed of these pieces has *a* as its topic. Thus, the link between identification and topic is not forged by any kind of cognitive item. An identification can be of x (because x is its source) and yet not contain information sufficient for distinguishing x from y.

The externalist fact applies only to perceptual identifications. In the case of a non-perceptual identification, the link between identification and topic is forged by means of fit or satisfaction of conditions. In such a case, the topic is the entity, if any, which fits the descriptive conditions stipulated by the body of information of the identification.

[16] The externalist fact can be rewritten in a number of different ways according to the theory to which one subscribes. For example, purely causal theories can propose that perceptual identification is of the entity that causes it.

If the resulting perceptual identification is composed exclusively of pieces of information about x, the identification has x as topic. It is a *univocal* identification. However, the bundling of pieces of information about x with pieces of information about y results in an identification that has x *and* y as topics. This results in an *equivocal* identification. Thus, I consider that aboutness of information is sufficient to explain what makes identifications univocal or equivocal.

> *The opacity fact*. Pieces of information about different things can be treated by the mind as being about the same thing and, further, different pieces of information about the same thing can be treated as being about different things.

Notice that the mere presentation of the opacity fact involves appealing to the externalist fact. The cognitive system is able to take pieces of information about distinct things as being about the same thing. Moreover, the cognitive system is able to treat pieces of information about the same thing as being about distinct things. As a result, the cognitive system is liable to insert information into inappropriate units of identification.

Let me recapitulate. Aboutness of information explains why the identification D is about x and not about y. Identification D is about x because D is composed exclusively of information about x. But the appeal to aboutness of information will not work for explaining how the mind is capable of treating pieces of information about the same thing as being about different things.

The opacity fact is due not to information but to the use of information by the cognitive system. This is the subject of the next section.

8 Supermaps

The connecting of pieces of information by the cognitive system can be classified into two types. *Vertical connecting* is the integration by the cognitive system of the pieces of information received in each particular episode of perception. In vertical connecting, the cognitive system uses only perceptually received information. Typically, perceptual identifications are the results of vertical connecting. *Horizontal* connectings are the different processes of interrelating stored

pieces of information which result in higher organization and integration of the informational body. In horizontal connecting, the decision as to whether information should be grouped as being about the same thing depends not only on perceptual clues but also on reasons, hypotheses and theories. In this case, the mind unifies pieces of information in terms of conferred aboutness. Reasons, hypotheses, and theories guide a lot of our inward working in conferring aboutness to pieces of information.

Vertical and horizontal connectings can happen in combination. In reidentification, for example, there is a combination of vertical and horizontal connectings.

What is horizontal connecting good for? What is its result?

We can say that the cognitive system has a subsystem – the *system of identification* – dedicated to identification and reidentification. The system of identification has as its ultimate goal the generation by means of vertical and horizontal connectings of a network of identifications capable of tracking the relevant uniformities of the world. In other words, the system of identification has as its aim the generation of the most complete picture of the world possible. I shall call it '*Supermap*'[17]. The formation of the supermap compels received information to fit together in appropriate ways to specify the relevant uniformities of the world.

We are detectors of information. But not only that - we are supermappers. Every piece of information throws light on its source and thus on fragments of situations of the world. We go beyond these registers of information. We integrate the parts into a coherent whole, filling the gaps, building hypothetical explanatory hierarchies, and so on. Supermaps are like representations of the physical and cultural world in which information is integrated and organized. Identifications are the dots in supermaps.

We can say, with some idealization, that the following principle rules over the cognitive system:

> Cartographic Principle (CP): The cognitive system works in order to form one identification for every relevant uniformity.

[17] The notion of supermap is inspired in the knowledge-map of Strawson (1974, 45-6).

There are three classes of cartographical correlations linking identifications and the entities tracked by them:

(a) Type 1-1 (univocal) – one identification for tracking x in its course in the environment;
(b) Type 2-1 (univocal but duplicate) – two (or more) identifications for tracking the same topic;
(c) Type 1-2 (equivocal) – one identification for tracking two (or more) entities.

The ideal shape for the supermap is the 1-1 type. 2-1 cases are cases of entries which seemingly are routes to different things and 1-2 cases are cases of equivocity. I shall call 'correct' the identifications which are in correlations of type 1-1.

Supermaps overwhelmingly constituted by correlations 2-1 and 1-2 are quite unfit to allow the agent to cope with the world because their dots do not match real divisions of the world.

The capacity to form horizontal connections is a good thing. It gives us supermaps. However, some errors are errors of horizontal connecting. In horizontal connecting, the agent can erroneously take a meeting with an unknown man as a reidentification of a friend or take his friend as another person. Further, the agent can take different properties or places as the same. In addition, the agent can take two occurrences of the same property as occurrences of different ones.

The cognitive system has the function of providing correct perceptual identifications by means of vertical integration of information received by the senses. This is part of the cognitive system's overall function of representing the environment by experience. When the cognitive system operates in a sound way, it is capable of forming correct perceptual clusters. Not to make the correct perceptual identifications is to dramatically lack fitness to the world.

Horizontal connectings are a different business. Such processes are more liable to produce equivocal or duplicate identifications even though they tend to yield identifications in cartographical correlations of type 1-1. The reason is that they are based not only on perceptually received information but also on beliefs, theories, hypotheses, and assumptions in order to establish the identity of the topics

of the identifications. The interference of beliefs and theories expands the possibilities of connecting items of information received in perceptually discontinuous episodes which really are about the same thing. However, this same fact exposes the cognitive system to the risk of formation of identifications in correlations of type 2-1 or 1-2.

The capacity to build horizontal connections increases our fitness to the world by making our provisions of information about uniformities more integrated and organized. The price for the use of this capacity is the increasing risk of error and formation of equivocal identifications. It is worthwhile.[18]

Breno Hax, Jr.
Federal University of Paraná

References

Adams, Fred. 1991. Causal Constraints. In *Dretske and His Critics*, ed. by B. McLaughin. Oxford: Basil Blackwell.
Barwise, Jon & Perry, John. 1999. *Situations and Attitudes*. Stanford: CSLI Publications.
Dretske, Fred. 1981. *Knowledge and the Flow of Information*. Cambridge/MA: MIT Press.
Dretske, Fred. 2000. *Perception, Knowledge, and Belief*. Cambridge: Cambridge University Press.
Evans, Gareth. 1982. *The Varieties of Reference*. Oxford: Clarendon Press.
Perry, John. 2002. *Identity, Personal Identity, and the Self*. Indianapolis: Hackett.
Strawson, P. F. 1959. *Individuals*. London, Methuen and Co.
Strawson, P. F. 1974. *Subject and Predicate in Logic and Grammar*. 2 Ed. Aldershot: Ashgate, 2004.
Woodfield, Andrew. 1991. Conceptions. *Mind* 100: 1991, 547-72.

[18] I am very grateful for the stimulating observations of Nathan Salmon, Marco Ruffino, Guido Imaguire and Luiz Arthur Pagani in the POL Lisbon Curitiba Philosophy Colloquium at the Universidade Federal do Parana at Curitiba in 2009, in which an earlier version of this paper was presented. I am also very grateful to Fred Dretske and to Eros Corazza for fruitful comments on an earlier draft.

On the Ontology of Relations

Guido Imaguire
Federal University of Rio de Janeiro and LanCog Group

BIBLID [0873-626X (2012) 34; pp. 689-711]

In *A Theory of Universals* (1978) Armstrong regretted that the interest around the dispute on internal and external relations had died away with the decline of Absolute Idealism. Further, in consequence of this, advances in the ontology of relations have not been comparable with the great advances in the topic of the logic of relations. Since Armstrong wrote this, some time has passed by, but the situation is essentially unchanged. Some investigation concerning relations has been done, but I think that many of them were contaminated by old prejudices.

This paper is intended as a contribution to minimize this gap. I will defend and argue in favour of some positive claims concerning the ontological status of relations. The main topics – and corresponding sections – of this paper are (1) irreducibility, (2) externality and (3) reality of relations. Thus, my topics exactly coincide with Russell's main topics in the theory of relations. But my purpose here is not an exegetical investigation of the controversy between classical authors like Russell, Bradley and Leibniz.[1] I will just offer some cursorily historical remarks in order to introduce the questions, but then I will propose some (I hope) original theses and arguments.

1 The irreducibility of relations

Aristotle was the first to defend the reducibility of relations. He said in his *Categories* that, whenever two (or more) substances are related, this is to be explained by means of certain monadic properties or

[1] Many scholars did such an analysis; see e.g. Clautenbaugh (1973), Mugnai (1992) and my own (2006).

accidents inhering in the relata.[2] Many scholastic and modern philosophers followed him in this. But, like Russell (and some medieval thinkers, in particular Albert the Great and John Duns Scotus), I believe this old tradition to be wrong. Some relations are reducible, some irreducible – that is my first thesis.

The topic of reducibility of relations has been usually treated as a semantical issue concerning relational propositions. Russell proposed its standard formulation: can all *aRb* propositions be reduced to *S-P* propositions? But the term "reduction" can also be interpreted in an (more original) ontological sense. And since my interest here does not concern semantics, I propose an ontological reformulation: do all relational facts supervene upon monadic facts? Actually, the semantical and the ontological questions must be not only distinguished, but clearly separated. Ontology does not always have to follow semantics. Indeed, after proposing a test for deciding which relations are ontologically reducible and which not, we will see that propositional reducibility does not imply ontological reducibility, nor the other way round.

Let us first see a misguided route. Many contemporary authors (e.g. Mugnai 1992: 96, Horstmann 1984) have suggested that relations are reducible if and only if they are symmetric. They are clearly influenced by Russell's argument[3]: if *R* in *aRb* were symmetric, than *R*{*a,b*} would be a correct analysis (the predicate "R" is applied to the set {a,b}). In this case we do not need the ordered pair (a,b) – the simple set {a,b} would do. A pair, insofar as it is ordered, is "implicitly relational", i.e. it entails a hidden relationality (something like "*a* comes before *b*") – so thought Russell and his followers.

Now, I think that this kind of reduction is a merely curious "logical" feature of symmetric relations. This might be relevant for a propositional reduction, but has no ontological relevance. (Actually,

[2] In *Metaphysics* Aristotle seems to relativize this doctrine claiming that "there are relational situations (such as Simmias's thinking about Socrates) in which substances are related not in virtue of a *pair* of accidents, but rather in virtue of a *single* accident possessed by just one of the substances." See *Stanford Encyclopedia of Philosophy* http://plato.stanford.edu/entries/relations-medieval.

[3] Russell used this argument many times, see e.g. *Principles of Mathematics* § 426, *The Philosophy of Leibniz* § 10, *Logical Atomism* p. 335, *My Philosophical Development* p.67.

I tend to think that this is more a notational than a real "logical" feature.) Take the case of Peter and his son Paul. Peter was first, say at t_1, taller than Paul. Along the years, Paul grew and became, say at t_3, taller than Peter. Since *is taller than* is an asymmetric relation, it must be irreducible according to Russell's criterion. But there is certainly a moment, say t_2, in which Peter and Paul had exactly the same height. Now, *is as tall as* is a symmetric relation, and, according to this criterion, it must be reducible. The consequence of this kind of reasoning would be that the particular height-relation between these two men changes its status through time: it is sometimes reducible (and so, according to Russell, internal and unreal) and sometimes irreducible (and so external and real)[4]. But this certainly sounds very odd for any philosopher with a robust sense of ontology. Paul's height relative to Peter (or any other object) can change through time, but the relation cannot change its ontological status (reducibility, reality or externality). This reasoning becomes even more absurd when we interpret – as usually is done – "internality" by means of "essentiality": Paul was at t_1 essentially (internally) smaller than Peter, then, at t_2, Paul became accidentally (externally) as tall as Peter, and, finally, at t_3 Paul became essentially taller than Peter. This is simply metaphysical non-sense. There cannot be diachronic change in the essential properties of things. Thus, I conclude that symmetry is not an adequate criterion for deciding the question of *ontological* reducibility.

Nevertheless, I think that Russell was right in defending that some relations are reducible, and others not. But the reason for this difference is not a curious logical feature like symmetry, but an ontological one. Rescher (1967:72) proposed a general and very useful test for deciding the question on the propositional reducibility of relations. In a simplified formulation, the semantical reducibility test (SRT) says:

> (SRT) The relation R, which subsists between two particulars a and b, is reducible if and only if there exist monadic predicates P_1, P_2, ..., P_n and Q_1, Q_2, ... , Q_n, such that $P_1(a) \wedge P_2(a) \wedge \ldots \wedge P_n(a) \wedge Q_1(b) \wedge Q_2(b) \wedge \ldots \wedge Q_n(b)$ determine completely the truth established

[4] For Russell, reducibility, externality and reality are (at least extensionally) equivalent.

by *aRb*.[5]

Two remarks are important here. First, the predicates P_1-P_n and Q_1-Q_n in the SRT should obviously not be explicitly or implicitly relational like "[…Rb]" (i.e. something like "to be taller than *b*"). When one accepts this kind of relational predicate the reduction of every relation becomes trivially possible. Thus, impure monadic predicates (predicates with an implicit reference to some particular) must be excluded. Second, ST is clearly dependent on the language we use: the "existence of predicates" is at stake. And the existence of monadic predicates depends, of course, on our arbitrary act of introducing such predicates in our language. But in this case our ontology would be trivially dependent on linguistic decisions. Thus, following the general device proposed above for separating ontological and semantical reduction, I suggest a purely ontological version of the reducibility test (ORT):

> (ORT) The relation *R*, which subsists between two particulars *a* and *b*, is reducible if and only if *a* has the properties P_1, P_2, …, P_n and *b* has the properties Q_1, Q_2, … , Q_n, such that the conjunctive fact $P_1(a) \wedge P_2(a) \wedge \ldots \wedge P_n(a) \wedge Q_1(b) \wedge Q_2(b) \wedge \ldots \wedge Q_n(b)$ completely determines the fact *aRb*, i.e. *aRb* supervenes the conjunctive fact with monadic properties.[6]

The basic insight of my ORT is the – I hope: reasonable – supposition that reducibility can be defined by means of supervenience. But note: I define reducibility by means of supervenience, but I do not identify reducibility with supervenience *simpliciter*. As we will see below, not all supervenient relations are also reducible. Again, two remarks are important here: First, similarly to our first remark

[5] Parkinson (1965: 45) suggest a stronger (and implausible) variant of this test, in which only the predicates of one leg of the relation - $P_1(a) \wedge P_2(a) \wedge \ldots \wedge P_n(a)$ – could sustain completely the truth established by *aRb*.

[6] I am not committed to – and will not argue for – the existence of complex facts in addition to simple facts. The phrase "complex fact $P_1(a) \wedge P_2(a) \wedge \ldots$" could simply be paraphrased by "the fact that $P_1(a)$ and the fact that $P_2(a)$ and the fact that…". Further, since this criterion also holds for not actually obtaining facts, the phrase "states of affairs" would probably be more adequate, but I will simply use "fact" for the sake of simplicity.

above concerning impure predicates, in the ORT impure monadic properties (implicitly relational properties) must be excluded in order to avoid trivial reduction. Second, in ORT I deliberately avoid talking in terms of "existence" or "non-existence" of properties, as was the case in SRT. The question on the existence of properties is connected with some more difficult metaphysical questions like the realism versus nominalism dispute, criteria for deciding which properties really exist (quantification over properties – by Quine – or their occurrence in natural laws – by Armstrong, etc.) that are to be avoided in this section. I will come back to this later. Therefore, I will simply take all ordinary properties to be genuine properties in the following discussion.

In general, we say that an entity B supervenes upon an entity A if and only if it is impossible that A should exist and B not exist, where A is possible[7] (this last restriction, of course, is introduced in order to avoid trivialization – for the rest of this paper I suppose it implicitly). For the particular case of relations, we can formulate:

> The relation *R* in *aRb* is supervenient when the relational fact *aRb* supervenes upon facts $F_1, \ldots, F_n$ if and only if it is impossible that $F_1, \ldots, F_n$ obtain and *aRb* does not obtain.

In this formulation, I say *R* "*in*" *aRb* is supervenient because we cannot suppose without more arguments that, given that the fact *aRb* is supervenient in our sense, the relation *R* is supervenient *simpliciter*, i.e. in all other facts where it occurs (it could be not supervenient in another fact *cRd*). Alternative standard formulations for supervenience will do as well for our purposes: "the facts $F_1, \ldots, F_n$ entail *aRb*", or "all $F_1, \ldots, F_n$-worlds are *aRb*-worlds", etc.

Now, how can we decide when a relation is supervenient in this sense? The first step here is the recognition of the huge variety of kinds of relations. Actually, I think that the recognition and observance of this variety is the most urgent methodological imperative in ontology of relations today. I am convinced that some philosophical theses about relations are misguided because of the failure in observing this variety. To put it in Wittgenstein's words "A main cause of philosophical disease – a one-sided diet: one nourishes one's think-

[7] This definition was proposed by Armstrong in (1997: 11)

ing with only one kind of example" (PI § 593). This kind of "one-sided diet" is specially striking when we compare the fanciless examples used in analyses of relations with the rich variety of examples used in analysis concerning particulars and monadic properties.

Relations vary in many different ways. For deciding the test of reducibility it is helpful to point out the variety of ontological "orders" (still in an ambiguous sense) of relations and introduce a notation for codifying these orders:

$R^{0\text{-}0}$ relations that obtain between objects
(e.g. *a* loves *b*, *a* sees *b*, *a* is bigger than *b*)
$R^{1\text{-}1}$ relations that obtain between 1-order properties of objects
(e.g. the wisdom of *a* is greater than the wisdom of *b*)
$R^{0\text{-}1}$ relations that obtain between an object and a 1-order property of an object
(e.g. *a* admires the wisdom of *b*)[8]
$R^{1\text{-}0}$ relations that obtain between a 1-order property of an object and an object
(e.g. the weight of this box surprised me)
$R^{2\text{-}2}$ relations that obtain between 2-order properties or 2-order relations of objects (e.g. the beauty of the colour of *a* is greater than the beauty of the colour of *b*).
$R^{0\text{-}2}$ relations that obtain between an object and a 2-order property or relation of objects (e.g. *a* admires the beauty of the colour of *b*).

and so on...

This classification is ambiguous: it allows a semantical and an ontological interpretation. I think that not all $R^{0\text{-}0}$ relations in a semantical sense are also first-order $R^{0\text{-}0}$ relations in a full ontological sense. Take e.g. "*a* is bigger than *b*": in a semantical sense, this is a $R^{0\text{-}0}$ relation: *R* has entities of type 0 as relata; "*a*" and "*b*" are singular terms. But, ontologically, the fact that *a* is bigger than *b* supervenes upon the fact that *a* has the height S1, and b has the height S2 (and S1>S2).

[8] We must distinguish between $R^{0\text{-}1}$ and $R^{1\text{-}0}$ because relations are not always symmetrical. Thus $R^{0\text{-}1}$ is a relation from an object to a property and $R^{1\text{-}0}$ a relation from a property to an object.

Again, language is not always a good guide to ontology. (There is a second ambiguity: my distinction of orders should not be confused with the Aristotelian distinction of orders – I will come back to this later.) In any case, the relevant interpretation for deciding reducibility is the ontological one. My point concerning reducibility is this: some relations supervene upon monadic properties of related things, others do not. Supervenient relations are ontologically dependent on the properties of things, and their holding or not is fully determined by the things having properties. Examples of supervenient relations are:

a is taller than b
a is as tall as b
a is taller than b
a is more beautiful than b
a is more famous than b

Note that, from a semantical point of view, these are all $R^{0\text{-}0}$ relations – but, and this is my point, not from an ontological point of view. These relations do not hold "directly" between particulars, but between properties of particulars. Take the first example again: the fact that *a* is taller than *b*. An ontological analysis of the corresponding fact should be something like: *a* has height *S1* ∧ *b* has height *S2* ∧ *S1>S2*. Thus, ontologically speaking, *is taller* is an $R^{1\text{-}1}$ relation. The fact that *a* is taller than *b* is entailed in the conjunctive fact that *a* is S1 ∧ *b* is S2 (e.g. there is no possible world in which *a* is 2m tall, *b* is 1m tall and the relational fact that *a is taller than b* does not hold).[9]

Of course, not all properties are objectively measurable. But this does not imply that, in these cases, the relation cannot supervene upon properties. Take e.g. beauty: if *a* is beautiful and *b* is ugly, then *a* is more beautiful than *b*. In such cases, objective analysis becomes difficult, but this does not undermine supervenience.

In all these cases, the particulars hold the corresponding relation

[9] Russell would certainly say, contrary to me, that this analysis shows that the relation *is taller than* is not reducible: in the *analysans* another relation (*S1>S2*) occurs, thus the relation did not "disappear". This makes it clear that my sense of "reducibility" does not correspond to Russell's sense.

just *because* they instantiate certain properties. The properties instantiated by the particulars necessitate the relation. The properties can be essential or accidental to the particulars, but they are a necessary and sufficient condition for the holding of the relation.

On the other hand, examples of non-supervenient relations are:

a loves b
a is married to b
a is the father of b
a kills b
a kisses b

These are $R^{0\text{-}0}$ relations from both semantical and ontological points of view. There is no property of *a* or *b* that entails the love of *a* for *b*. Maybe *a* loves some property of *b*, but this is a different fact from *a loves b* (in that case, we would have $R^{0\text{-}1}$ instead of $R^{0\text{-}0}$). Maybe *a* loves *b because* of some property of *b*, but this is, again, another (I suppose complex) fact: it is a fact that explains why *a* loves *b*. Maybe *a* is married to *b* only because *b* has the property of *being rich*, but this does not mean that it is no property of *a* which is married to any property of *b*, but *a is married to b*. Note that some relations are dependent on some properties: *a* is father of *b* only if *a* instantiates the property *to be male*, *a* can kill *b* only if *b* instantiates the property *to be alive*, etc. In these cases, the properties are necessary conditions for the relations, but they do not necessitate the relation. A male does not have to be father, a living organism does not have to be killed, and so on. The same is valid for all levels of relations. Take the $R^{0\text{-}1}$ relation

a admires the wisdom of *b*
a envies the wealth of *b*

These are clearly non-supervenient relations: there is neither a property of *a* nor a property of *b* that necessitates the relation of admiration (not even *b*'s wisdom). Similarly, there is neither a property of *a* nor a property of *b* that necessitates the relation of envy (not even *b*'s wealth). A relation is reducible if and only if both sides are grounded in properties of the terms, i.e. the relation must be "both-sided" supervenient.

All this sound as if reducibility of relations were identified with supervenience *simpliciter*, but this is not the case. In the proposed test for reducibility, it is required that *aRb* supervene upon the complex fact with *monadic* properties, while in the definition of supervenience for relations, it is not required that facts $F_1, \ldots, F_n$ must be monadic. This is so because it seems possible that a given relational fact F_1 supervenes upon two facts F_2 and F_3, such that one of them (or both) is relational and irreducible. Take e.g. the relation *is-brother-of*. I think the relational fact *a is brother of b* supervenes other relational facts (*a* is a child of x with y ∧ *b* is a child of x with y), and these last relational facts are not supervenient and irreducible. If this is correct, *is brother of* is a case of a supervenient but irreducible relation, i.e. it is not reducible to exclusively monadic facts. Therefore, reducibility and supervenience, at least for relations, are not equivalent. One might ask: if this is true, how should we decide that a relation is reducible or irreducible until the "whole analysis" to ultimate basic facts is done?

Indeed, the question on a complete analysis (of facts, of propositions, or whatever) seems to lead us into deep trouble. How can we decide whether at a given moment we achieved the terminal level of the most basic facts? The proposed criterion for deciding reducibility of relations seems to be committed to the assumption that facts are not infinitely analysable. Actually, I believe this to be a plausible assumption. The possibility of an infinite analysis displeases most philosophers. Of course, there is, insofar as I can see, no knock-down argument against it. Thus, it seems that a relational fact F_1 can be (completely) reduced to a conjunction of exclusively monadic facts $F_1, \ldots, F_n$, but that some (at least one) of these facts, in their turn, supervene on even more basic relational facts, and so on. In this case, a conclusive decision about reducibility of relations seems unachievable. But even if there were such infinite complex facts, my ORT offers a conclusive criterion at least for the positive cases, i.e. that a certain relational fact is reducible. Note that ORT does not require that the constituents of the complex fact $P_1(a) \wedge P_2(a) \wedge \ldots \wedge P_n(a) \wedge Q_1(b) \wedge Q_2(b) \wedge \ldots \wedge Q_n(b)$, that necessitate the fact *aRb*, must be ultimate. For a given relation *R* to be reducible, it suffices that at "some" level a full eliminative conjunction of monadic facts is achievable. In this case, we can conclude that *R* is reducible. What we cannot conclude, of course, is that all relations are reducible (in a linguistic

formulation: we cannot conclude that a complete monadic "world-description" is possible). But, again, by the plausible assumption that analysis is not infinite, we have a definitive criterion for both cases of reducibility and irreducibility.

It is important to stress that this ontological distinction between reducible and irreducible relations does not coincide with the Aristotelian distinction between first order and second order attributes.[10] To say that a relation is grounded on the properties of things does not imply that this relation is, in the Aristotelian sense, a second-order relation. When Socrates and Plato are white, the fact that Socrates and Plato have the same colour entails the reducible relation *to have the same colour*. But I think Aristotle would (correctly) consider *to have the same colour* (different from *is the same colour as*!) a relation between things (and not between properties of things). Reducible relations in my sense are, first, ontologically dependent on properties of things, i.e. the properties are a necessary condition for the supervenient relation. Second, the relations are always present when the corresponding properties are instantiated in the relata, i.e. the instantiation of the relevant properties is a sufficient condition for the subsistence of the supervenient relation. These relations are, to use Armstrong's terminology (1997: 12), an "ontological free lunch" – they are not an addition of Being.

Of course, it is not easy to decide for every relation whether it is supervenient to other facts or not: is the relation *is-20m-distant-from* a non-supervenient relation or a relation that supervenes upon the monadic properties *position-of-x* and *position-of-y*? This depends, of course, on our conception of space. Suitable ontological analyses are required in many particular cases; in particular, spatial and temporal relations deserve a special examination.

The very natural thesis defended here has been overlooked by many authors simply because they restrict their analysis to a too small set of examples of relations, instead of analysing their great variety. Mugnai e.g. (1992: 22) comes very near to my criterion. He tells us that for Leibniz *veritates relationum videntur esse reflexivae*, i.e.

> truths corresponding to relations are properly speaking 'second-order

[10] This could be one reason why Aristotle was unable to recognize different degrees of reality of relations.

> truths', which, so to say, supervene other truths. The relationship of similarity between Socrates and Plato, for example, happens at the movement when Socrates and Plato are both white, but this simply means that the truth of the proposition: 'Socrates is similar to Plato' is based on the truth of the two assertions: 'Socrates is white' and 'Plato is white', and the whole proposition comes into being when one reflects on these assertions.

Of course, the truth of "Socrates is similar to Plato" is grounded on the truth of "Socrates is white" and "Plato is white". Thus, this relation can be considered reducible. But on which monadic propositions is the truth of "Socrates loves Plato" based? I cannot see any monadic propositions and no distribution of monadic properties in the world that could ground its truth.

2 On the internality of relations

One classical dispute between the pioneers of analytic philosophy and the absolute idealists concerned the internality vs. externality of relations. Absolute idealists, at least according to the standard interpretation, defended that all relations are internal (e.g. F. H. Bradley in 1897), while analytic philosophers (in particular Russell) defended that some relations are external. But it is not clear what these authors mean by the phrases "internal" and "external". In *My Philosophical Development* (1959) Russell gave us some hints about how he understood the thesis of internalism of relations:

> He [Leibniz] says that, if a man living in Europe has a wife in India and the wife dies without his knowing it, the man undergoes an intrinsic change at the moment of her death. This is the kind of doctrine that I was combating. (MPD: 42)

But it is also not clear how we should interpret the vague word "intrinsic" in this passage. Russell could hardly be combating the principle of identity according to which entities with different properties are different. If *x* is a husband (and not a widow), and *y* is a widow (and not a husband), then it is necessarily true that *x* is different from *y*. It seems very plausible to suppose that the poor man of the story changed a property with the death of his wife. And Russell would certainly not deny that the wife in the story undergoes an intrinsic change at the moment of her death. In contemporary theory of properties, it is usual to distinguish between intrinsic and extrinsic prop-

erties. Intrinsic properties are non-extrinsic, and extrinsic properties are properties that a thing has in virtue of its relation to another (at least one other) thing, i.e. they are implicitly relational – in the example, *to be a husband* and *to be a widow* are certainly extrinsic. But this distinction was not explicitly made by Russell. And if this were what Russell meant, his claim would be trivial – of course, the change of the man was a change in virtue of something else, namely the death of his wife. Thus, his intuition must be based on a distinction between "somehow" more relevant and less relevant properties and relations. The most obvious distinction of this sort is the classical distinction between essential and accidental properties. Actually, in some passages Russell explicitly introduced the notion of "essence":

> "If A and B are related in a certain way", it may be said, "you must admit that if they were not so related they would be other than they are, and that consequently there must be something in them which is *essential* to their being related as they are." (MDP: 46, my emphasis)

In any case, I will discuss in this section the status of relations concerning their essentiality. Since Russell was not a friend of the notion of substance and essentialism, I suppose he would not be happy about this decision. His usage of "essential" in this passage is probably only a *façon de parler*. But this should not worry us – our aim is systematic and not exegetical. After all, this was certainly Moore's original usage. In his "External and Internal Relations" (1919) he characterized internal relations modally: internal relations are essential to their relata.

I will defend two theses in this section:

(i) there are some (at least "impure" or "supervenient") essential relations, and
(ii) the question of essentiality (internality) is independent of the question on reducibility or irreducibility of relations (at least in the case that origin essentialism is correct).

Thesis (i) is, once one accepted essential properties, in some sense trivial. If there are essential properties at all, this is a difficult question I cannot investigate here. I will simply presuppose some kind of orthodox (Kripkean) essentialism in the following. Thesis (ii) is a more substantive thesis. It is intended as a criticism of Russell's (and

others') supposition that reducibility is equivalent to internality (as well irreducibility to externality).

The overwhelming majority of examples of essential properties discussed in literature are monadic and not polyadic. As far as I can see, the most relevant case of an essential relation is the origin essentialism suggested by Kripke (1972: 112f), according to which if *a* is the mother of *b*, *b* is essentially son of *a* (no person could have another mother than he or she actually has). And it is not surprising that this example of an essential feature is much more controversial than its monadic rivals. It is more intuitive to suppose that the existence of an entity depends on one of its monadic "constitutive" properties, than on something "external" as a relation that this entity holds with another entity. A "free floating" Socrates in an empty universe without any relations to other things (even without his mother) may sound strange, but less strange than a non-human Socrates. Concerning origin essentialism Robertson (1998) points out that both the origin essentialist of an artefact (that this table should be made of this very hunk of matter) and the origin essentialist of living organisms (that this organism must be developed from this very collection of propagules) are in trouble for they cannot respect their own essentialist intuition that a limited variation of the origin of a given thing is possible. It both cases, origin essentialism seems to be based on mereological intuitions: one thing cannot be construed of completely different constituent parts. The egg and the sperm are built into the new organism as essential parts of it. In the following I will suppose origin essentialism to be correct.

Now, when one is willing to assume a more tolerant view about what relations exist –i.e. not denying that "impure" or supervenient relations are genuine relations – we obtain an easy strategy for defending the existence of essential relations. We could simply construct essential relations from essential (monadic) properties: when *to be a human being* is an essential property of both Socrates and Plato, as most think they are, then *to belong to the same species as* is an "impure" essential relation between them. I call this relation "impure" because it entails an implicit reference to a monadic property. Actually, the essential character of this relation is only derived from a monadic property. One may object that this relation is not really essential, since one of them, say, Plato, could have not existed. In this

case, Socrates would not belong to the same natural kind as Plato. I think this to be clearly wrong, because Plato is essentially a human being – independent of his existence. On the contrary, his very existence depends on being human. In all possible worlds in which he exists, he is human, and in all possible worlds in which he does not exist, he is, in some sense, human too. It is not the case that, in these possible worlds in which he does not exist, he is something else, say a worm. Thus, independent of his existence or not, independent of which possible world we take, Socrates always belongs to the same natural kind as Plato.

Whether such artificially constructed relations are respectable entities depends on how tolerant we are in our ontology. Of course, these relations are, according to my own criterion reducible, insofar as they are simply grounded in monadic properties. Just as most philosophers are not willing to accept that every predicate express a genuine property, so many may reject such relations as *to belong to the same natural kind as*. But at least (i) can be so far concluded: there are impure reducible essential relations.

Let us now think about the relation between essentiality and irreducibility. Some contemporary authors are already persuaded by Russell and think that internality (here: essentiality) is the same as (at least coextensive with) reducibility. See for example this passage by Heil (2009: 315):

> Following the medieval thinkers, we could say that internal relations are "founded" on monadic (non-relational) features of their relata: when an internal relation holds, it holds in virtue of non-relational features of whatever it relates.

And, concerning external relations, he continues

> An external relation, R, holds between objects *a* and *b*, in virtue of *a*'s being F and *b*'s being G (F and G being non-relational properties of *a* and *b*). [...] Further, the initial breezy characterization of internal relations, can now be seen to apply quite generally: if you have the relata, you have the relations. (Heil 2009: 317)

I think this (in particular the general form "quite generally") to be clearly wrong: there is neither a necessary connection between internality (essentiality) and irreducibility nor between reducibility (to be "founded" on monadic properties) and externality of relations. Heil's failure to see the independence of both aspects lies probably

in the meagre diet of examples he took, namely the internality of six being greater than five and the external monadic founded fact of Simmias being taller than Socrates. I am convinced that *is greater than* in *six is greater than five* is internal not because it is grounded in monadic properties of five and six, but simply because pure arithmetical properties and relations are always essential (or internal). And the relation *is taller than* in *Simmias is taller than Socrates* is external not because it is not founded on monadic properties, but because these monadic properties (the height of Simmias and the height of Socrates) are not essential. Thus, I propose that, when *a* is essentially F and *b* is essentially G and the relation R supervenes F and G, then R is internal to *a* and *b* (this was e.g. the case of *belong to the same species*). When *a* is not essentially F or *b* is not essentially G (or both), then the relation R that supervenes F and G is external to *a* and *b*.

Of course, first one has to obtain clarity about which relations are really essential (if any). Let us, for the sake of the argument, suppose that origin essentialism is correct. Thus, *a is son of b* is an internal relation (to *a*, not to *b*). Remember that thesis (ii) has a hypothetical form: when origin essentialism is correct, the question of essentiality (internality) is independent of the question on reducibility or irreducibility of relations.

The easier strategy for arguing for the independency of reducibility and internality is to give some counter-examples. A very plausible example of an accidental relation is *to love*. It can be hardly denied that although *a* actually loves *b*, *a* could not simply love *b* without damage to its existence or identity. Nevertheless, *to love* is, according to our ORT, an irreducible relation. When *a loves b*, this relation is not supervenient to any property of *a* or b. Thus, there are irreducible accidental relations. Irreducibility does not imply internality.

Take now the relation *is heavier than* in *a is heavier than b* (where *a* and *b* are, say, human beings). It is clear that this relation can be reduced to monadic properties in the sense of ORT (*a weighs x Kg* ∧ *b weighs y Kg* ∧ *x*>*y*). Weight is not an essential property of a person, except, or course, for super-essentialism, for which every property is trivially essential. This is an example of a reducible accidental relation; thus, reducibility does not imply internality.

According to ORT, the relation *to be son of* is irreducible, for if *a is son of b*, this is so not in virtue of some monadic properties of *a* and

b. Now, the relation *is son of* (to have these very parents) is considered an essential property of a person. Thus, there are internal irreducible relations; internality does not imply reducibility.

Take, on the other hand, another example of Kripke's (1972: 113-114) for an essential property, namely *to be made of*. According to him, if this table is made of wood, we cannot say that *this very table* could, instead of wood, simply be made of ice water taken from the River Thames. A table made of ice from the river would not be *this* table. If this is correct, then the relation *is made of the same substance as* is an essential relation for two objects *a* and *b* (suppose, made from wood). Nevertheless, given our ORT, *is made of the same substance as* is a reducible relation, since the fact that *table a is made of the same substance as table b* supervenes upon the monadic facts that *a is made of wood* and *b is made of wood*. Thus, there are relations that are both internal and reducible; internality does not imply irreducibility. To sum up, as (ii) claims: internality/externality and reducibility/irreducibility are independent ontological aspects of relations.

3 Reality of relations

In contrast to "existence", "subsistence" and "actuality", no definitions have been offered for the notions of "reality" and "unreality". It seems unclear what we mean when we say that something is real or unreal. In most cases, we use this notion in ontology in a more or less intuitive sense. Further, these terms seem not to qualify something in a positive, interesting way: Are there real and unreal things? I propose in this section two different approaches to the problem of reality, the traditional and the contemporary a posteriori one.

3.1 The traditional approach

According to this approach, reality is conceived by means of ontological independence. The more an entity is ontologically dependent on others, the more it is unreal. This approach has a long tradition. In scholaticism God was considered the *ens realissimum* just because He is absolutely independent of any other entity; Aristotelian substances were considered real (even if not "as" real as God) because they were independent of their attributes; these attributes, in their turn, were

less real than substances because they were dependent on them, and so on. Fictions are ontologically weak because they are dependent on human creative thought. Following this reasoning, some scholastics conclude that relations have a very low ontological status, for their existence relies on the existence of many other entities: on the relata, on the properties of these relata and, worse, on a recognizing mind. Thus, when *Plato and Socrates are similar* (both are white), this similarity depends on Plato, Socrates, the whiteness of Plato, the whiteness of Socrates, and even on a mind that compares both.[11]

But some medieval thinkers tried to see a little bit more reality in relations. Ockham for example made a distinction between *relationes reales* and *relationes rationis*. While *relationes rationis* are dependent on an arbitrary decision of our intellect, like the relation between a word-sign and its meaning, *relationes reales* depend on the recognition of an intellect, but not on our arbitrary decision, like the relation between a mother and her son. Thus, Ockham did certainly not intend to reject the objective nature of the *relationes reales*. And when Leibniz says that relations are *entia rationis*, he neither intends them to be arbitrary creations of our intellect, like Ockham's *entia rationis*, nor to be mere fictional entities. Leibniz distinguished between *entia rationis sine fundamento in re,* like Pegasus, and *entia rationis cum fundamento in re*, like relations that subsist in actual facts. If David is father of Solomon, then the relation of paternity between David and Solomon is not a mere fiction in the sense that Pegasus is one.

But to say that relations have a *fundamentum in re* can still be considered a sign of ontological weakness: relations do not subsist in themselves. They need the relata in order to subsist—they are grounded *in re*. Indeed, in his letter to Des Bosses (17 fev 1711) Leibniz says that a relation without *fundamentum* is a self-contradictory notion. Relations are not ontologically independent and do not have independent "capacity of movement". They begin to subsist and end subsisting, but only insofar as their "ontological basis"—objects or properties of objects—allow or determine this. Thus, the thesis of

[11] In *De Potentia*, q.8 ar. 1, ra 4 Thomas writes: "Relatio habet esse debilissimum, quod est eius tantum" and very similarly, so does Scotus in *Super Praedicamenta* (q. 25, 10): "Relatio inter omnia entia est debilissimum ens, cum est sola habitudo duorum, et ita minime est cognoscibilis in se" (quoted by Mugnai 1992:27)

unreality of relations should be understood in the sense that relations are ontologically dependent, and not that they lack an objective nature.[12] This insight seems to be present in natural intuition: Imagine a collection of many particulars organized in a hierarchical order by the ordering relation *R*. Now what would remain if we subtract all the particular things? An "empty order", an ordered structure without ordered things, is non-sense.

Now, what should we conclude? I do not think that all relations are mind-dependent. Most late scholars defended the objective nature of some relations, and this seems a plausible thesis. A knockdown argument against mind-dependence of any kind of entity cannot be given, nor for the particular case of relations. The best we can do, I suppose, is to follow contemporary metaphysicians and take natural science with all its relational concepts (causality, cross fertilization, chemical reactions, etc.) in order to see how objective relational descriptions of reality are. If we also take mathematics, the question seems even more evident: relations are not eliminable and, thus, objective. Of course, a radical idealist would not be impressed by this argument, but I suppose him to be unimpressed by any realist argument at all. The need of relational predicates does not imply the existence of an ontological correlate.

In any case, when my argument in the first section is sound, there are relations that are irreducible to monadic properties. When *aRb* is a relational fact that does not supervene upon more basic monadic facts, then we should conclude that *R* is not ontologically dependent on monadic properties. At least these relations are not an "ontological free lunch". Therefore, at least irreducible relations must have the same ontological status as their monadic rivals, the properties. Of course, one could even suppose that relations are independent of concrete instantiations, like Platonists suggest for monadic properties, but this would led us to a too fundamental problem of philosophy – the debate between Platonism and nominalism – which I

[12] Another classical argument against the reality of relations is Bradley's famous argument of regress. If the relation R which holds between *a* and *b* were real, we would need another relation, say R*, which holds between this real entity R and a, such as a relation between R and b. But if R* is a very real entity again, R* need also a new relation, say R**, which holds between R* and a, like as between R* and R, and so on *ad infinitum*.

prefer to avoid here.

One could suggest a further tentative argument for reality of relations. As we said above, entities are more real the more independent they are. In addition, one could also plausibly argue that an entity is more real, the more entities depend on it. Thus, God would be an *ens realissimum* not only because He is independent, but also because all other entities depend on Him. Now, when origin essentialism is correct, one could conclude that, when *a is the son of b*, *a* could not exist without the relation *is son of b*. Thus, ex hypothesis, *a* is ontologically dependent on this relation. In this case, of course, dependence would be symmetric, and both entities (*a* and its relation to *b*) to the same degree real.

Many, including Russell, thought that supervenience is a criterion for unreality. But this seems to be a prejudice. To say that heat is nothing else than motion of molecules does indeed commit one to say that heat is ontologically dependent on molecules, but not to say that heat is unreal. It is interesting to note that we tend naturally to think this relation of dependence as asymmetrical (heat depends on molecules, but not conversely), although the existence of temperature is coextensive with the existence of molecules: there cannot be molecules without some motion nor molecules without any temperature (repose is a case of motion just as 0°K is also a temperature). Similarly, as soon as things exist, so do relations exist, too: at least relations like self-identity or difference, but also relations like *as heavy as*, *as great as*. When the objects are qualitatively different (which is required by the Leibnizian principle of identity), there must also be some asymmetrical relations (*is heavier than*, *is smaller than*, ...). Thus, the set of all possible worlds in which things exist, coincides with the set of all possible worlds in which relations exist. Thus, in the modal interpretation of the notion of dependence, relations and things are mutually dependent and so must have the same degree of reality.

3.2 The a posteriori approach of ontological commitment

With the expression " *a posteriori* approach" I do not mean a single doctrine, but a family of approaches that have in common the contemporary gloss of delegating to total science a central role for de-

ciding questions of existence. The two most prominent theories are, insofar as I can see, Quine's criterion for ontological commitment and Armstrong's (and Schoemaker's) a posteriori Realism.

With this approach, we learned to discuss questions of *existence* instead of *reality* of things. After all, existence is nearly a technical term. Of course, there is no agreement about it: is it a first or a second order predicate? Is there a difference between existence and subsistence? Is ontological existence nothing more than what is expressed by the quantifier? In any case, with this more technical term, philosophers now have the feeling they are making some progress on these questions.

According to Quine's famous suggestion, ontological questions should be decided by means of the ontological commitment test. The test consists on a simple procedure: to translate the sentences of our best theories into canonical logic, determine the domain of quantification required to render this translation true, and read the entity commitments of the elements of the required domain. Of course, some paraphrasing is allowed during the translation – and here lies the clue of the story. Oriented by these rules, many philosophers engaged themselves in the Quinean ontological game: (1) searching for sentences of natural science containing quantification over predicates (the Platonists), (2) trying to paraphrase these sentences without quantification over predicates (the nominalists), (3) showing that these paraphrases were not correct (the Platonists again), and so on. Thus, when we are willing to engage in this game for deciding the existence of relations, we should simply search within our best theories for sentences containing quantification over relational predicates.

The second contemporary criterion for "existence" or "reality" was proposed by Armstrong (1978). According to his "a posteriori Realism", total science has to decide about what universals there are. Like Quine, he thinks that the mere existence of predicates is not sufficient to support the conclusion that a corresponding universal must exist. And, again like Quine, science plays a decisive role in deciding the existence of (monadic or dyadic) properties. But Armstrong's criterion is not based on the determination of the domain of quantification of theories, but the question of what properties are linked in natural laws. Thus, for the particular case of relations, the

question would simply be: are there non-monadic properties linked by natural laws?

I think that the observance of some very simple examples would suffice to show the necessity of the commitment to relations for both criteria, the Quinean and the Armstrongian. Take, e.g.:

(1) Some physical interactions between spatially distant bodies are only attractive (gravity), others both attractive and repelling (electromagnetism).
(2) Some causal connections in medicine (substances over physiological processes) are explained, others unexplained.
(3) Some alloying of metals can generate a material that is stronger than their primary elements. (e.g. steel is stronger than iron)
(4) There is probabilistic (non-deterministic) causality in quantum mechanics.
(5) Some matrimonies are recognized only in some, but not in all countries (e.g. homosexual matrimony).
(6) Some killing is morally and legally justified (e.g. in case of self-defense)

I think all these sentences to be good candidates for supporting the existence of relations. They are considered true by our total world description and (some) are linked by natural laws (others by social sciences). Moreover, they all involve quantification over a relational predicate, and no paraphrase without quantification over the relation seems available. Of course, in each case, a detailed logical analysis would be required. But even if there were such a paraphrase, a search in the field of science would yield many others candidates.

Take (1) for instance. By existence generalization we can derive that there are physical interactions between spatially distant bodies, and interaction is essentially a relation. Take (5). The marriage of two men *a* and *b* conducted in a country where homosexual matrimonies are legal is not recognized in another country. This does not mean that the particulars *a* or *b* are not officially recognized, but that *a* is married to *b* (the relation between *a* and *b*) is not recognized. Similar analysis could be offered for the other examples. In any case, I think by observing our normal world descriptions we obtain

enough grounds for supposing that relations are both constituents of natural laws (as Armstrong requires) and values of bounded variables of true sentences (as Quine requests).

4 Conclusion and open questions

At first sight, this paper could be seen as a defense of Russell's position, when I argued for irreducibility, externality and reality of—not of all, but of any—relations. Nevertheless, my ontological approach contrasts radically with Russell's logical approach. For Russell, relations are external and real *because* they are irreducible, i.e. irreducibility, reality and externality are equivalent features of relations. I argued here, in contrast, in favour of the thesis that relations can be external and real independently of being irreducible.

One important methodological conclusion I think we should draw from this analysis is that philosophers are all too fast in concluding general claims based on too few examples. In Wittgenstein's words "A main cause of philosophical disease – a one-sided diet: one nourishes one's thinking with only one kind of example" (PI § 593). When I am right, some philosophers (including Leibniz, Mugnai and Heil) failed to correctly understand relations due to nourishing their thinking with only one kind of examples.

Guido Imaguire
Federal University of Rio de Janeiro and LanCog Group

References

Brower, J. 2009. Medieval Theories of Relation. In Stanford Encyclopedia of Philosophy. http://plato.stanford.edu/search/searcher.py?query=medieval+relations

Burkhardt, H. Anmerkungen zur Logik, Ontologie und Semantik bei Leibniz. *Studia Leibniziana*, Band VI, Heft 1, Franz Steiner, Wiesbaden, 1974.

Clatterbaugh, K.C. 1973. Leibniz's Doctrine of Individual Accidents. *Studia Leibniziana*, 4, Wiesbaden.

Cover, J.A. & O'Leary-Hawthorne. 1999. *Substance and Individuation in Leibniz*. Cambridge: Cambridge Univ. Press.

Hartshorne, C. 1970. *Creative Synthesis and Philosophical Method.* London: The SCM Press.

Horstmann, R.P. *Ontologie und Relationen*: *Hegel, Bradley, Russell und die Kontroverse über die interne und externe Beziehungen*. Königstein, Ts: Athenäum, 1984.

Imaguire, G. 2006. A Critica de Russell à Concepção Leibniziana das Relações. *Manuscrito*, v. 29, n. 1.

Krikpe, S. 1972. *Naming and Necessity*. Oxford: Basis Blackwell.
Leibniz, G.W. *Monadologie*.
Leibniz, G.W. *Metaphysische Abhandlung*, ed. by Herrings. Hamburg: Meiner, 1991.
Leibniz, G.W. *Generales Inquisitiones de Analysi Notionum et Veritatum*, ed. by Franz Schupp, Hamburg: Felix Meiner, 1982.
Leibniz, G.W. *Die Philosophischen Schriften* (GP), ed. by C.I. Gerhardt. Berlin, 1857-90. Reprint: Hildesheim: Olms, 1965.
Leibniz, G.W. *Nouveaux Essais*. English translation *New Essais on Human Understandings*, transl. and ed. by P. Remnant and J. Bennett. Cambridge: Cambridge Univ. Press, 1981.
Linsky, B. *Russell's Metaphysical Logic*. Stanford: CSLI Publications.
Linsky, B. The Metaphysics of Logical Atomism. In *The Cambridge Companion to Bertrand Russell*, ed. by N. Griffin. Cambridge: Cambridge University Press, 2003.
Moore, G. 1919. External and Internal Relations. *Proceeding of the Aristotelian Society* 20:40-62. Reprint in Philosophical Studies, London: K. Paul, Trench, Trubner & Co: New York: Harcourt, Brace & Co., 1992, pp. 276-309.
Mugnai, M. 1992. *Leibniz' Theory of Relations*. Studia Leibnitiana Supplementa. Stuttgart: Franz Steiner Verlag.
Mulligan, K. 1998. Relations – Through Thick and Thin. *Erkenntnis* 48: 325-353.
Parkinson, G.H.R. 1965. *Logic and Reality in Leibniz's Metahysics*. Oxford: Oxford Univ. Press.
Rescher, N. 1967. *The Philosophy of Leibniz*. New York, Englewood Cliffs.
Russell, B. *An Analysis of Mathematical Reasoning*. 1898. In *Collected Papers 2.*
Russell, B. *The Philosophy of Logical Atomism* [PLA] 1918. In *Logic and Knowledge*.
Russell, B. *Analytic Realism* [AR] 1911. In *Collected Papers 3.*
Russell, B. *The Basis of Realism* [BR] 1911. In *Collected Papers 3.*
Russell, B. *My Philosophical Development* [MPD] 1959. London: Routledge, 1993.
Russell, B. *The Principles of Mathematics* [PoM] 1903. London: Routledge, 1992.
Russell, B. *Logical Atomism* [LA] 1924. In *Logic and knowledge*. London: George Allen and Unwin, 1956.
Russell, B. *A Critical Exposition of the Philosophy of Leibniz* [PL] 1900. London: Routledge, 1992.
Russell, B. The Nature of Truth [TNT] 1905. In *Collected Papers 4*.
Sleight, R. C, Jr. 1900. *Leibniz and Arnauld: A Commentary on Their Correspondence*. New Haven: Yale University Press, 1990.
Sprigge, T. L. S. 1973. *The Vindication of Absolute Idealism*. Edinburgh: Edinburgh University Press.
Wittgenstein, L. 1953. *Philosophical Investigations*. Ed. by G.E.M. Anscombe and R. Rhees. Transl. by G.E.M. Anscombe. Oxford: Blackwell.

Spontaneous Linguistic Understanding: a few Introductory Remarks

André Leclerc
Federal University of Ceará-Fortaleza

BIBLID [0873-626X (2012) 34; pp. 713-737]

Let me try to explain from the outset what I mean by "spontaneous linguistic understanding" (SLU). As a first approximation, it is the non-reflexive, fluent, direct and mostly non-inferential process by which a speaker-hearer *dynamically*, by considering usually stretches of discourse rather than isolated sentences, determines the *intuitive* truth conditions (or, in general, the satisfaction conditions) of the *utterances* performed in a specific context of use. Most people on Earth enjoy that kind of experience all the time in their mother tongue. The view of SLU I shall try to develop here takes actions and plans as a starting point, that is, utterances usually performed as part of bigger plans. Rarely do we perform isolated actions; our actions, as a rule, are parts of plans, steps towards the achievement of a project. This, of course, holds for sequences of speech acts in a discourse or conversation. As there are always different ways to perform the same act-type, it is no surprise that there may be different ways to perform any speech act-type. [1] Utterances are actions, and SLU, as we shall see, is tied to occurrent understanding of utterances. SLU is also a dynamic process that considers a topic that can be developed in different ways. However, it is obvious that what has been said in a discourse or a conversation smoothly restricts what can be relevantly said soon afterwards. In this process, truth conditions are grasped, especially those *actually* grasped and considered by the

[1] An act-type is denoted by an infinitive verb or a gerund; speech acts types denoted by illocutionary verbs share the same property: their performances vary and produce slightly, but meaningful different effects usually perceived by the hearers.

agents of the context (intuitive truth conditions), and not those that derive only from the knowledge of the general conventions and rules of language.[2]

Perhaps, the best way to delimitate the foregoing concept is to contrast what it represents, first, with "interpretation" (by that I mean hermeneutic practices, the most common and the most sophisticated as well), clearly a reflexive and inferential activity, and second, with cases of communication when one does not master completely the language of the addressee. Then I present the distinction dispositional/occurrent understanding and two famous philosophical programs: I see Davidson's program as a hypothesis about the structure of dispositional understanding, while Gricean program provide an interesting view about utterance (occurrent) understanding. I show that both programs are inadequate. I close the paper with a few suggestions on the status of semantic knowledge, on the immediate object of SLU, and an attempt to defend the conjecture that spontaneous linguistic understanding is grounded on the more basic understanding of actions and situations.

1 A preliminary delimitation of the concept

A) Some historical milestones

As usual in contemporary philosophy, the idea is not radically new. SLU was a great concern in the tradition of ideational theory of language, especially as developed by the most important proponent of that classical approach in the XVIIth century: Antoine Arnauld. He established such a contrast between the way most people, most of the time, understand utterances and judge the meaning of words, and the work done by hermeneutists. This way of judging spontaneously the meanings of words he called "*sentiment*," something comparable to Chomsky's speaker intuition, but not limited to syntax. It is, he said, the most universal and common way of judging almost everything. It is also "the surest, the finest, and the subtlest."

[2] On intuitive truth conditions, see François Recanati, *Truth-conditional Pragmatics*, Oxford, O.U.P., 2010; also his *Literal Meaning*, Cambridge, C.U.P., 2004.

> This is the way human beings assess almost all the variety of things in the world. We recognize at once that two very resembling persons are nonetheless different, without paying attention to details, to what is in the face of one that is not in the face of the other. The impression marks all this in the mind, without revealing distinctly the particular differences.[3]

Arnauld believed that this is the usual form of judgment we apply when we judge instantaneously (or "feel") that two words with very similar meanings have nonetheless different senses or conditions of application. The *sentiment* is a basic discriminatory capacity; applied to semantics, it allows us to discern "at once the finest differences between expressions better than all the rules in the world."[4] The mind just "feels" differences that cannot be made explicit without difficulty. In these cases, as Michael Polanyi nicely once put it, "we can know more that we can tell". Interestingly, Arnauld says that "the rules themselves are true only whenever they conform to the *sentiment*."

Spontaneous understanding is the basis or starting point of any hermeneutic work. Arnauld contrasted this "*sentiment*" with the pig-headedness of some hermeneutists (invariably his enemies, the Protestants) who searched for hidden senses, sometimes a bit *recherché*, behind the words pronounced by ordinary people in ordinary circumstances. For Arnauld & Nicole, to talk is to excite intentionally ideas in the mind of the hearer, or to cause intentionally a global impression in the mind of the hearer, and that impression is usually composed of ideas articulated in the sentence used, ideas inferred from the clues given by the speaker, and ideas neither articulated nor inferred (mimics, tone of voice, expressive character of gaze, etc.). So there is a clear distinction between the immediate understandings of utterances, that is, the global and complex impression received in a context of use, and the result of a discussion about what we *should*

[3] Antoine Arnauld & Pierre Nicole, [1669-1672], *La Grande Perpétuité de la foi de l'Eglise catholique sur l'Eucharistie*, [*G.P.*], publiée par l'Abbé M***, Paris, Imprimerie de Migne, chez l'éditeur rue d'Amboise, Hors la barrière d'Enfer, 1841, Vol. 2, Book 1, p. 990. My translation. See also my paper (Leclerc, 2005). M. Polanyi, *The Tacit Dimension*, Chicago, The University of Chicago Press, 1966/2009, first chapter, "Tacit knowing", where Polanyi made very similar remarks.

[4] Ibid, Vol. 2, Book II, chap. 1, p.122.

understand in the same context.

More recently, Wittgenstein changed the focus in his own work from the theory of meaning to the theory of understanding, and discussed at some length "immediate understanding" (*unmittelbar Verstehen*), from the *Big Typescript* until the end of his life. As we know, linguistic understanding became a main concern in Wittgenstein's last philosophy. In contrast, the *Tractatus* does not pay much attention to linguistic understanding, with the notable exception of T.4.024, which establishes for the first time an analytic connection between the notion of linguistic understanding and that of truth-conditions. In the *Philosophical Investigations*, Wittgenstein strives to show that understanding is not a state, an event, a process or an experience. As a matter of fact, there is no single experience (or state, process, etc.) to which linguistic understanding can be reduced. This is not to say that there are no experiences at all involved in the understanding of a sentence or an utterance. Michael Dummett insisted, rightly, that "we need an occurrent sense of 'understand'…"[5] Otherwise, the understanding of an utterance would be quite mysterious.

Still more recently, Burge drew a distinction between "comprehension" and "interpretation," the first being basically non-reflexive and non-inferential:

> Comprehension is understanding that is epistemically immediate, unreasoned, and non-inferential. First-person comprehension is the minimal understanding presupposed in any thinking, in beings that understand their thoughts at all.
>
> […] I include words, in a derivative sense, as things one can comprehend in the first-person way. One comprehends the words in one's idiolect as one uses them. The comprehended words are the *direct* expression of thoughts one comprehends. They express one's thoughts without mediation of further words or thoughts.
>
> […] Interpretation arises out of there being a question or issue about how to understand a candidate object of interpretation. Interpretation is always from the third person point of view. I conjecture that it is

[5] Dummett, M. (1993). *Origins of Analytical Philosophy*. Cambridge (MA), Harvard University Press, p. 60; also on p. 103. For a discussion, see Guy Longworth, "A Plea for Understanding", in S. Sawyer (ed.), *New Waves in the Philosophy of Language*. Aldershot: Ashgate, 2009.

always epistemically inferential.[6]

Interpretation is needed when something strange or surprising comes out, when there is a conflict or a disagreement as to the meaning of a sentence or of a stretch of a discourse. Interestingly, according to Burge, the first instance of understanding is the understanding of our own thoughts, and they are understood, *mainly*, in a non-inferential way. Of course, there are exceptions. After all, sometimes, people get confused; they don't know exactly what they want, for instance. And there can be degrees of understanding, or an incomplete grasping of a proposition. One may discover that one's belief that P has some unexpected presupposition or consequences. This corresponds to a deepening of our understanding. Thoughts readily expressible by linguistic means (full-fledged thoughts) and expressed by others are understood immediately when there is no need to interpret them.

B) Linguistic understanding in a language not fully mastered

Now suppose you are a tourist visiting a country speaking a language you do not master very well. Setting aside segmentation problems, you can understand, at least on some occasions, what the natives *mean*, but the understanding of what they *say* exactly is a painful and frustrating process. You have to pay attention constantly at every word in every sentence; you have the impression that the natives speak very fast all the time; and to form a less than secure interpretation of an utterance, you have to make a lot of inferences based on analytical hypotheses – to fill the gaps for the words you do not know yet – and on contextual clues. At night, you're back to the hotel, usually with a headache. After a few weeks, you return to your homeland and speak with the members of your family. Here is the contrast: At home, you enjoy fluent, effortless experiences of linguistic understanding. Very much like perception, these experiences are almost passive. You continuously get an "automatic," fast and direct access to intuitive satisfaction conditions for any sentence

[6] Tyler Burge, "Comprehension and Interpretation", in L.E. Hahn (ed.): *The Philosophy of Donald Davidson*. Chicago and La Salle, Open Court, 1999, 236-237.

of any syntactic type. This is what I call "spontaneous linguistic understanding."

2 Semantic knowledge

The analytic tradition has it that the understanding of a sentence is a kind of knowledge (knowledge of truth-conditions or satisfaction conditions).[7] However, if that knowledge is propositional knowledge, we can raise serious doubts on that traditional tenet. Is understanding a kind of knowledge at all? Some epistemological issues here are unavoidable.[8] I believe *there are* experiences of understanding, even if these experiences are semantically irrelevant to determine the meaning of the word "understanding," or even if none of these experiences could be correlated with something called "understanding." They are epistemologically important in order to distinguish linguistic understanding from the mere knowledge of the content of an utterance (What Is Said), and I believe they play an important role in the epistemology of testimony. SLU can be compare to perception on that score. I am allowed to testify in a court because I saw the murderer shooting the victim at midday, at an appropriate distance with my perfect vision, etc. In the same way, I am allowed to testify that, next door at the hotel, I heard a male voice in a tone of menace saying "I'm gonna kill you, bitch!" and then a female voice screaming "Help!". When the judge ask me, "Are you sure you heard just these words?" I can answer something like: "Well, the walls of the hotel are thin, the sounds was pretty distinct, and the accused speaks my mother tongue." The experience of hearing the sounds and the experience of direct discourse recognition enable me to testify. The knowledge of what has been said (the content of an utterance) can also be obtained through a reliable translation, but understanding requires the autonomous exercise of conceptual abilities, the use of our own semantic knowledge. So the kind of autonomous direct

[7] This has been challenged recently. See, Dean Pettit (2002), "Why Knowledge is Unnecessary for Understanding Language", *Mind*, Vol. 111, n. 443, July 2002, 519-550.

[8] Understanding is one of the main topics in the Epistemology of Language. See Alex Barber (ed.) (2003), *Epistemology of Language*, Oxford, O.U.P., 2003.

access to content that characterizes understanding (in contrast with indirect access through translation) presupposes at least these experiences.

If understanding is knowledge at all, what kind of knowledge is this? The Epistemic View takes understanding to be propositional knowledge of meaning. There are two steps leading to that conclusion.[9]

1. A understands S iff A knows what S means;
2. A knows what S means iff A knows *that* S means *m*.

∴ A understands S iff A knows that S means *m*.[10]

Step 2 is highly questionable. According to the Epistemic View, knowledge of meaning is always propositional knowledge. If propositional knowledge of meaning is analyzed in terms of true justified belief (or any variation on that classical analysis, that is, one that involves a belief), we should ask what kind of belief is that? The basic axioms of a Davidsonian T-Theory do not qualify to be the content of a genuine belief. Disquotational axioms like

I. "Piaf" refers to Piaf;
II. "Ella" refers to Ella;

or a compositional axiom of the form

III. "a is more famous than b" is true iff the denotation of "a" is more famous than the denotation of "b",

are not suited to be the content of a genuine belief, as Gareth Evans showed convincingly. A genuine belief can serve different projects. If I believe that the water in the bottle on my desk has been poisoned (Evans' example), I can do different things on the basis of that belief: committing suicide, killing someone by offering the water or by preparing a drink with that water, pouring the content of the

[9] See Dean Pettit (2002).

[10] "A" stands for any agent of the context; "S" for any expression or sentence; and "*m*" for the meaning of S.

bottle into the sink in order to prevent a tragedy, calling the police for investigating the case, etc. But what can I do with the belief that, say, "Lisbon" refers to Lisbon? [11] If the basic axioms of a T-Theory *à la* Davidson do not qualify as content for genuine belief, and if belief is a necessary ingredient in any analysis of propositional knowledge, then it seems that semantic knowledge cannot be propositional knowledge. Therefore, knowing what S means is not the same as knowing *that* S means *m* (where *m* can be a proposition, a concept, an idea, a function, a rule or an object).

Furthermore, propositional knowledge is "gettierable," that is, it fails in Gettier cases. When a justified belief represents a contingent-historical fact, it is always possible to devise a case *à la* Gettier for the corresponding knowledge. But this does not work for linguistic understanding. Suppose that knowledge of meaning is propositional knowledge. Imagine you just immigrate in Brazil and start learning Portuguese with a closed group of friendly people. After a while, you talk good Portuguese fluently with these people. But they have enemies and these nasty, mischievous enemies of your new friends convince you, in English, that the Portuguese the friendly people taught you is all wrong; worst, they used false evidences that your friends misled you intentionally on most linguistic matters relative to Portuguese. Furious, you leave the community (the friendly people) believing that they played a trick on you and that you have good reasons not to thrust what you have "learned" with them. You now believe you haven't learned good Portuguese and you decided to start over learning by yourself in the street. Surprisingly, however, you have the impression that you understand perfectly what any pedestrian you encounter randomly tells you; moreover, their behavior confirms your understanding. In such a predicament, it seems that firm propositional knowledge is not necessary for understanding. The impression that you are understanding already counts as understanding. Here is one of Pettit conclusions:

> I went on to argue that understanding language does not even require belief. That is, to understand a bit of language with a certain meaning,

[11] Gareth Evans, "Semantic Theory and Tacit Knowledge", in S. Holtzmann & C. Leich (eds.) *Wittgenstein: To Follow a Rule*, London, Routledge, 1981. Also in G. Evans, *Collected Papers*, Oxford, Clarendon Press, 1985/2002, 322-342.

> it is not necessary even to believe that it has that meaning. It is sufficient, I have argued, that it *seem* to you to have that meaning, whether you believe it or not.[12]

There is some resemblance between Pettit's conclusion and the *sentiment* described by Arnauld & Nicole, that spontaneous judgment on meaning on which the rules are based. Be that as it may, it is commonsensical that we know what we have learned (and not forgotten), and that natural languages are among the things we must learn. It is also obvious that in order to be able to understand a language, we must possess encyclopedic knowledge about the world, about the natural and social regularities. Linguistic conventions are social regularities. As it is hard to characterize satisfactorily the semantic knowledge needed to talk and understand normally a language, I shall adopt, for the time being, a very modest approach. I shall avoid any intellectualistic view, including the epistemic view, and make the following, naïve presupposition: SLU presupposes a huge set of dispositions, some enabling discourse recognition, while others, activated simultaneously, determine the contribution of sub-sentencial parts to the satisfaction conditions of illocutions of any type. By doing this, I just follow uncompromisingly Evans' suggestion:

> I suggest that we construe the claim that someone tacitly knows a theory of meaning as ascribing to that person a set of dispositions—one corresponding to each of the expressions for which the theory provides a distinct axiom.[13]

I think, however, that we must associate two dispositions to each expression. It is possible to recognize the token of a word as being the token of a type without knowing its meaning. Of course, Evans was interested only in semantic knowledge, but for those interested in SLU, discourse recognition is fundamental. The theory of meaning he has in mind is one that adopts the format advocated by Davidson in "Truth and Meaning" (1967), with basic axioms for each expression, singular terms and predicates as well, from which are derived theorems (the famous "T-sentences") displaying the truth-conditions for each declarative sentence of the language. Interestingly, Davidson himself recognized that when he was speaking of "the theory a hear-

[12] Pettit, *op. cit.*, 548.

[13] G. Evans, *op. cit.*, p. 328.

er has when he understands a speaker", this was a mere *façon de parler*.

> [...] I do not speak of implicit knowledge here or elsewhere: the point is not that speaker or hearer has a theory, but that they speak and understand in accord with a theory—a theory that is needed only when we want to describe their abilities and performance.[14]

That amounts to a kind of instrumentalism in theory of meaning. This is wise and prudent. But I think more can be said about the *structure* of dispositional understanding. I now turn to the distinction introduced by Michael Dummett between *dispositional understanding* and *occurrent understanding*. I shall defend that dispositional understanding must be relatively stable, while occurrent understanding is context-sensitive.[15]

3 Dispositional and occurrent understanding

While the theory of meaning has always been a main concern in analytic philosophy, the theory of linguistic understanding, strangely enough, did not receive the same attention. But we already have at our disposal a few useful distinctions, and two important philosophical programs to start with.

As a first step, we have to distinguish two basic kinds of linguistic understanding. Firstly, the "occurrent understanding" of utterances; in that case what we understand is what the *speaker* means in the context of utterance. The understanding of utterances is not only the understanding of sentences-token; it is also and more basically the understanding of actions performed for such and such a primary reason in a highly specific context. Secondly, we also have a "dispositional understanding" of sentences, expressions, bits of language; in that case, we understand what *sentences* mean, and they mean what they do in virtue of conventions, that is, social regularities of a certain type. The first kind of understanding clearly depends on the second kind of understanding. Quick occurrent understanding presupposes the existence of a huge set of dispositions acquired along

[14] D. Davidson, "The Social Aspects of Language", in *Truth, Language and History*, Oxford, Clarendon Press, 2005, p.113.

[15] For the distinction, see Michael Dummett, *Origins of Analytical Philosophy*, *op. cit*.

the first years of a child's life (in the case of a mother tongue). Segmentation and discourse recognition would be impossible without this set of dispositions. This set represents the knowledge we have of a language (at least of our idiolect), that is, the abilities to speak and understand, to write and read. [16] Basically, this is the knowledge of sound patterns (phonological knowledge) and the knowledge of what these sound patterns are regularly used to mean (semantic knowledge). Consequently, our concept of spontaneous linguistic understanding is tied to that of occurrent understanding. Dispositional understanding is not "spontaneous". Occurrent understanding is the autonomous exercise of an ability that always takes place in a specific context of utterance.

Dispositional understanding (the understanding of language or bits of language) is the tacit knowledge of the sound pattern and meaning an expression has as a *type*. It is the knowledge a competent speaker-hearer brings with him/her in any new context of utterance. There is an answer by "yes," "no," or "a little" to questions like: Do you know Japanese? Do you know Javanese? Do you know Spanish? By answering "yes," you are saying that you got through a process of learning, that you acquired and now possess a big set of dispositions that enables you to associate "automatically" senses to the characteristic sounds of a language. I think Austin's concept of *descriptive conventions* is quite useful here. A descriptive convention correlates a word (a categorem) to *kinds* of things (objects, states of affairs, situations). Lexical meaning in natural language is specified by a "descriptive convention". What we call "meaning," in this sense, is abstracted from social regularities. But the descriptive conventions only specify *kinds* of objects, kinds of situations and states of affairs or facts. Take the word "coffee". There is no such thing as "coffee-in-general". The descriptive convention for "coffee" must include, mainly, coffee beans, coffee powder, and coffee in a liquid state. Lexicographers have the choice between writing many different dictionary entries for the same word (a bad choice), or just a few

[16] For an interesting suggestion as to the structure and working of these dispositions, see Gareth Evans, "Semantic Theory and Tacit Knowledge", *op. cit.* For critiques and discussion of Evans' ideas, see also Crispin Wright, "Theories of Meaning and Speakers' Knowledge", in *Realism, Meaning and Truth*. Oxford: Blackwell, 1993, 204-238.

entries, indicating how different meanings derive from a main "core meaning." The second way of doing lexicography is certainly recommendable. It follows Grice's Razor in semantics: Do not multiply meanings beyond necessity, and keep semantics as simple as possible! For the word "coffee," when it is used to refer to the color, this use is clearly derived, and the same holds for "a coffee" (a cup of coffee), a kind of metonymy. The "core meaning" first described in a dictionary is something like "the seeds of a tropical bush from which, once roasted and crushed, a black, bitter and stimulating drink is made by adding hot water". Sometimes the stress is on the drink and not the seeds, but the whole information must be there. Determined by what Austin called "demonstrative conventions" (that correlate words and sentences to historical, real, specific objects and situations), related but slightly different semantic values are derived in context when "coffee," for instance, is used in an utterance to refer to coffee in a specific state (fresh coffee, old bitter coffee, coffee beans, etc.).

Tacit knowledge of meaning is what feeds the experiences of understanding, but it is not itself an experience, at least not in the same sense as sensations, perceptions, emotions, imaginations or memories are said to be experiences (or parts of experiences). Tacit knowledge of a meaning is a disposition. These "automatic" sound-sense associations are practically independent of the will, like perception. As George Lakoff points out, it is impossible to follow the command: "Don't think of an elephant!" [17] When we hear the word "elephant" we cannot help but thinking of an elephant, of a large animal with floppy ears and a trunk. We do not choose the meaning of the words we understand as we do not choose to see what we see when we open our eyes. Of course, there are exceptions, "local agreements" on meanings and stipulative definitions, but it is obvious that we cannot do that massively and all the time. *So, descriptive conventions, the object of dispositional understanding, must be relatively stable.* Otherwise, it would be possible for everyone to develop something like a private language, so that, at the end of the day, our idiolects would be too divergent for the sake of communication.

Are there dispositional understandings of tokens produced on an

[17] See George Lakoff, *Don't Think of an Elephant!* White River Junction (Vermont), Chelsea Green Publishing, 2004.

occasion? Well, suppose you open a text book of Basic English Grammar and read a sentence given just as an example of a grammatically correct sentence, say "John and Mary are going to school by bus". You do understand something, but this poor understanding of an isolated sentence does not give you the knowledge of any specific situation, because the tokens of the words simply inherit the semantic properties of their corresponding types. No new and richer semantic values are contextually derived and the sentence, consequently, is not clearly truth-evaluable. The understanding we have of the sentence certainly does not qualify as a case of spontaneous linguistic understanding. There are no intuitive truth conditions easily associated with the sentence. This is clearly not a serious and literal use of language. In semantics, since Frege, we all adopt the convention according to which the assessment of word meaning or a sentence meaning must proceed only on the supposition that the word or sentence is used seriously and literally, and, I would like to add, *as part of a whole discourse*. The interpretation of isolated sentences, normally, is not much more than a simple conjecture. Sometimes a sentence is used seriously and literally but the result is not very convincing. Take the sentence "There is life on Mars" (Dummett's example). If the understanding of a sentence is the knowledge of what is the case if it is true, what kind of knowledge do I have in such a case? [18] The mere disquotational truth conditions for that sentence ("There is life on Mars" is true iff there is life on Mars) are totally uninformative. An exobiologist would certainly have a different understanding from the rest of English speakers. Are there occurrent understanding of standing meaning (for expressions-type)? Well, it happens each time you take a dictionary, but once again, it does not qualify as "spontaneous."

Standing meaning, by definition, must be relatively stable. Nonetheless, in natural languages, new uses of old words appear constant-

[18] For a different opinion, see Guy Longworth, "A Plea for Understanding", in Sawyer, S. (org.). *New Waves in the Philosophy of Language.* Aldershot: Ashgate, 2009, 138-163. As I said, cases of dispositional understanding of standing meaning of expression-type, certainly do not qualify as "spontaneous." Occurrent understanding of standing meaning emerge precisely when we do not understand the sentence used on a first reading; we have to stop, to think twice and to analyze its structure, check the lexical meaning, etc.

ly. Take the word "here". "Wait for me here!" usually means "around here", not too far from the spot where you are right now, that is, where the utterance takes place. But nowadays, in many web pages, you find instructions saying: "Click here!", and to follow the command, you have to click on the very token of the word "here". This is a new (funny) use of the word, but no one ever had any problem in understanding it. Of course, the new use is related to the previous ones. Words in natural languages don't have by convention a limited number of uses or possible understandings or senses, and our linguistic policy does not determine in advance all the correct understandings of a term.

The occurrent understanding of utterances usually involved *tokens* produced in the context of utterance. We always produce these tokens in a *specific* situation. There is no such thing as coffee-in-general. The coffee we refer to is always fully determined, in a specific state (liquid, powder, or beans, etc.). Here the Austinian "demonstrative conventions" do the job, close the gap between the general and the specific, so to speak. A full semantic value for the token is derived in the context. Let me expand a little bit an interesting example given by J. M. Moravcsik.[19] The word "walk" in the sentence "Jones had a walk" means different things if Jones is a healthy adult (she walked a few kilometers to stay in shape), or a toddler (she just gave her first steps in her whole life), or an elderly person in a hospital recovering from a serious disease (she walked painfully from his bed to the bathroom and back), or an athlete that undergone a surgery in his knee (she will be back to her team soon), or if Jones is always seen running (by contrast, exceptionally, she walked!). In natural languages, most of the time, a sentence is correctly understood when it is understood differently in different contexts. *The upshot here is that the spontaneous linguistic understanding of utterances in natural languages is context-sensitive.* [20] *And the first and immediate object of SLU is occasion meaning*. It is useful to compare natural languages with the regimented languages

[19] See Julius Moravcsik (1998), *Meaning, Creativity and the Partial Inscrutability of the Human Mind*, Stanford, CSLI Publications, 1998.

[20] That understanding in natural language is context-sensitive is a thesis developed at length by Charles Travis in *Unshadowed Thought*, Cambridge (MA), Harvard University Press, 2000. See also *Occasion-Sensitivity*, Oxford, O.U.P., 2008.

used in sciences, where a sentence is correctly understood when all the members of the scientific community understand it in the same way. For that reason, ordinary language is unsuited for systematic theorizing, or for scientific investigation and communication. The betterment of scientific communication is precisely what justifies regimentation. In a regimented or ideal language, special (technical, theoretical) words and concepts are introduced through definitions, and to define is precisely to delimitate conditions of application.

Occurrent understanding of sentences proceeds, first, by identifying the derived semantic value of the sub-sentencial parts of the sentence; this always precedes any act of predication. The next step consists in identifying the act of predication, I mean, what is predicated of what. This is the same as grasping a proposition. When the proposition grasped is a general one or a complex one, the identification of the first and second order predication is required; otherwise, the argument must also be identified in case of atomic proposition of the form F(a). In case of relations, especially asymmetric relations, the order of predication is of course important. Predication (or functional application) is the cement of propositions. The identification of various constituents and structural elements looks like a complex process, but this is just a product of analysis. When I see an old friend in a crowd, I do not pay any special attention to her eyebrows, chicks, hairs, the color of her eyes, etc. All this is familiar, and it is enough: I recognize her at once. The same holds for the spontaneous understanding of a sentence: if I know the words, and if the structure is not too complex, the proposition (or truth conditions) is grasped at once.

4 Davidson's programme

How do we have a so quick access to the content of an utterance when the token produced is the token of a sentence never heard before? Davidson's famous answer in 1967, before the qualifications added a few decades later, is that we must master somehow a compositional T-theory for the language we use. I see Davidson's conception as an attempt to describe the structure of our dispositional understanding. The T-theory is a theory of meaning (and understanding) for that language, and he argued that the format of such a theory is analogous

to that of a formal system. First, we have a huge set of basic axioms specifying: a) the reference (meaning) of singular terms and b) compositional axioms for n-ary predicates; second, the T-sentences or theorems derived from the basic axioms and specifying the truth conditions for all the declarative sentences of the language. For instance, from the basic axioms I), II) and III) we can derive the following theorem or T-sentence:

IV. "Piaf is more famous than Ella" is true in English iff Piaf is more famous than Ella.

The theorem gives the truth conditions of the sentence-type mentioned, and explained how someone could understand that sentence on a first hearing.

There are a few worthy problems to be mentioned here, especially for those who believe that knowledge of meaning (as specified by the axioms) is propositional knowledge, and those who "psychologized" Davidson's program in a cognitivist vein. 1) Truth conditions are not enough. As David Wiggins points out, when you try to pick up the right truth conditions of a sentence, say, "The Sun is behind cloud" (Wiggins' example), you must already have an understanding of the sentence in order to stop the list of many other things you might put on the right side of the corresponding T-sentence: "The Sun is behind cloud" is true in English iff the Sun is behind the cloud, *and* it is day time, the sun has risen, there are more people awake than asleep, etc. [21] Furthermore, we understand illocutionary acts, and not only propositional clause, that is, a pair composed of a force and a propositional content, and any strategy to reduce the non-declarative to the declarative fragment of a natural language faces problems, especially in the case of expressive illocutionary acts. 2) An informative and non-circular specification of our knowledge of meaning expressed by the basic axioms has still to be produced; without a mode of presentation for the meaning, it seems that that cannot be done (this is

[21] See David Wiggins, "Meaning and Truth Conditions: From Frege's Grand Design to Davidson's", in B. Hale and C. Wright (Eds.), *A Companion to the philosophy of Language*, Oxford, Blackwell, 1997, p. 7. For similar remarks, see Scott Soames, *What Is Meaning?*, Princeton, Princeton University Press, 2010, chapter 3.

Stephen Schiffer's mode of presentation problem).[22] 3) As we have seen, a genuine propositional attitude is one that may serve many purposes and interact with other attitudes, as Gareth Evans rightly observed. This is clearly not the case of the knowledge of meaning expressed by the basic axioms. So it is at least doubtful that the basic axioms express genuine propositional knowledge. Almost 20 years after "Truth and Meaning" (1967), Davidson acknowledged the fact that the mastering of a T-theory is not enough for the understanding of an utterance and that its specification is always incomplete; an interpreter also needs a "passing theory".[23]

5 Grice's program

Grice's program focuses on occurrent understanding, that is, the understanding of utterances. The understanding of an utterance takes place when the hearer (the audience) recognizes the speaker's intention to cause a certain effect precisely by the recognition of that intention. The whole process of identification (recognition) of speaker's intention is regularly taken to be an *inferential* process. Grice's program faces problems too. 1) It does not accommodate very well the conventional aspects of communication and understanding; 2) It does not cope quite well with situations of counter-suggestion, confession and the anti-lying problem;[24] 3) Above all, it describes occurrent understanding mainly, if not exclusively, as an inferential process, so that there is no special difference between the understanding people have when they are speaking their mother tongue, and the

[22] See S. Schiffer, "The Mode of Presentation Problem", in C. Anthony Anderson & J. Owens (Eds.), *Propositional Attitudes. The Role of Content in Logic, Language and Mind.* Stanford, CSLI Publications, 1990, 249-268. Also, by the same author, "Knowledge of Meaning", in A. Barber (Ed.), *Epistemology of Language*, Oxford, O.U.P., 2003, 303-324.

[23] See D. Davidson, "A Nice Derangement of Epitaphs", in Lepore, E. (ed.), *Truth and Interpretation: Perspective on the Philosophy of Donald Davidson*, Oxford, Basil Blackwell, 1986.

[24] For an exposition of these problems, see Alex Barber, "Truth Conditions and Their Recognition", in A. Barber (ed.), *Epistemology of Language*, *op. cit*, 367-395.

kind of understanding a tourist has in a foreign country. Moreover, Grice's approach does not represent correctly the difference between the spontaneous linguistic understanding of ordinary people in ordinary circumstances, and the work done by hermeneutists. In other words, the Gricean program does not describe correctly occurrent, fluent, effortless understanding. For that reason, like Millikan, Recanati and others, I believe that understanding is better modeled as a kind of perception.[25]

6 Perception and inference: a digression

Sometimes the difference between perception and inference is not that clear and it seems that we have a mere difference of degree. You need to change your spectacles and go to consult your oculist. There are the letters projected on the wall; the first lines are easy, but the letters become smaller and smaller until you reach a point at which you are not anymore clearly perceiving, nor clearly inferring: Is it an "E" or an "F"?

In Leclerc (2009, 267-268) I mention

> [...] a common distinction in the literature about two exclusive ways of representing communication and linguistic understanding: Firstly, there was the Code Model, sometimes called "the Expressive View", associated with Aristotle's *De interpretatione* and also to Port-Royal's *Grammaire générale et raisonnée*. This is certainly and by far the most enduring and influential conception of human communication in the whole history of philosophy. [...] The aim of any act of communication according to the Code Model is to share the same meanings. Secondly, there is the Inferential Model, sometimes called the "Convergence View," initiated half a century ago by Grice and developed also by Sperber & Wilson. Davidson (1986) can also be seen as a proponent of this view. Here, the measure of success in human communication is the correct identification of speaker's meaning or intention through inferential processes, or to converge on the same meanings and to make the same inferences.

That Code-Inference opposition in the theory of communication and understanding is one more idealized opposition in which no one fits,

[25] Ruth Millikan, *Varieties of Meaning. The 2002 Jean Nicod Lectures*, Cambridge (MA), MIT Press, 2005; and *Language: A Biological Model*, Oxford, O.U.P., 2005. Also, François Recanati, "Does Linguistic Communication Rest on Inference?", *Mind & Language*, Vol. 17, numbers1 and 2 February/April, 105–126, 2002.

one more procrustean bed. Both models are clearly incomplete. SLU must accommodate short and fast inferences. The stock example is the sentence "She took the keys and opened the door". We all understand (infer) that she opened the door *with the keys*, even if that component is not articulated in the sentence. A mother says to her son who is crying because of a very small cut on the tip of a finger: "Come on, you are not going to die" (Recanati's example); the intuitive proposition (or truth conditions) expressed here is that the boy in not going to die *because of the small cut*, and not, of course, that the boy is immortal. Many cases of contextual "enrichment" (Recanati's expression) are of this kind. The same holds for malapropisms. But not everything is inferred in linguistic understanding.

We have seen that there is an intuitive difference between knowing the content of an utterance, and understanding it. If a reliable translator tells me that an utterance in Russian means that Putin is wise, I am in a position to claim that I know the content of the utterance, but certainly not in a position to claim that I understand the utterance. SLU requires a strong degree of epistemic intimacy with the tokens produced in the context and the autonomous exercise of tacit semantic knowledge. To use a distinction made in Phenomenology and Gestalt psychology, in the perception of the tokens produced in the context of utterance, the subsidiary attention is directed to the sounds perceived (discourse recognition), but the focal attention is directed to the senses of the expressions, to the content of the utterance. In the same way, the pianist's focal attention is directed to the melody and the quality of her interpretation, not to the position of her fingers. Similarly, when we perceive objects around us, the focal attention is directed to the objects themselves and not to the shape, size or colors.

Grice's program is based on Mindreading (the identification of speaker's intention is certainly a form of Mindreading). But I don't think that Mindreading is best reconstructed as a kind of inference by analogy. It can be reconstructed as a kind of perception too. This is how I understand Alvin Goldman's simulation theory of Mindreading. Mindreading has been invoked a long time ago as a cornerstone of language use by Antoine Arnauld, classified as a proponent of the Code Model. Here is a quote from the Great Arnauld:

> One cannot reflect, however little, on the nature of human language without recognizing that it is entirely grounded on that imperfect penetration of the mind of the others. That is why, when speaking, there are many things we do not need to express. (My translation)

We wouldn't speak the way we do if our minds would be totally opaque to each other.

There is another analogy between perception and SLU. Perception is a dynamic process based on a retention-projection structure. We do not perceive all the sides of an object at once. On the basis of what we just perceived we project or anticipate the perception of the other sides. Sometimes we commit mistakes. We thought that we were seeing a house, but it is only a façade put there by the staff of a filmmaker. But perception is a self-correcting process. Misperception is corrected by perception. SLU is also a dynamic process. Conversations usually involve turn-taking and sequences of utterances. Rarely do we have to understand isolated utterances. The utterances we just understood determine what Grice called in his famous Principle of Cooperation "the accepted purpose or direction of the talk exchange in which you are engaged." [26] They also help us to restrict the range of possible interpretations for the new utterances produced in the talk exchange and to project further possible developments in the conversation.

7 Conclusion: plans, scripts and situations [27]

The approach I am trying to develop is not new. It is a bottom-top approach in which it is semantics that must answer to pragmatics, as Brandom once put it. The approach is based on ideas that can be traced back to the works of Arnauld, Peirce, Wittgenstein, Austin and more recently Travis and Recanati. In Leclerc (2009, 271), I gave it the following wording:

[26] Paul Grice, "Logic and Conversation", in P. Grice, *Studies in the Way of Words*. Cambridge (MA), Harvard University Press, 1989, p. 26.

[27] Part of the material of this section has been published in *Manuscrito*, Vol. 32, nº1, jan-jun, 2009, Special Issue: *Semantic Content and Communication*. A. Leclerc, E. Perini-Santos & M. Ruffino (eds.). See my "Meanings, Actions and Agreements", in that issue, 249-288.

> Actions and practices are the roots of linguistic understanding. In a pragmatic approach, we start with actions, of which speech acts are important sub-species, and actions are what we interpret. Actions are performed for different reasons, and are parts of bigger plans. *Words and sentences are seen and understood as instruments used in the performance of actions.* ***My suggestion is that we should consider plans or scripts and not only actions, as the unit of investigation***. I also suggest that the expectations of agents (speakers-hearers), and the correct identification and understanding of these expectations, are especially important for the determination of the sense of the words used in an occasion. In other words: ***The root of any distinction in thought and in the sense of linguistic expressions is found in its sensible effects, in our practices, plans and activities.*** This principle I call the Principle of the Determination of Sense.

What makes this principle so important is the phenomenon of the plasticity or underdetermination of sense in ordinary language, the fact that different tokens of the same sentence-type may have different truth conditions in different contexts of use. On that score, Austin's theory of truth accommodates that phenomenon much better than Tarski's.

SLU presupposes a great dose of "worldliness and reasonableness", as Travis (2000) used to say. We share the same world, as Davidson pointed out. However, taken at its face value, that does not take us very far. The knowledge of the world is very different from one person to the next. And the world we live in – not the one described by physics – is forever changing. The same could be said of the knowledge of our language (as the norm of a community). We know that social class, education and learning, scientific discipline, influences from good or bad neighborhood, etc. may contribute to create discrepancies among the idiolects of the people living in the same community (What we call "ordinary language" itself is changing all the time, even if you define it as the degree zero of "regimentation".)

But "our knowledge of the world" can be understood alternatively as something that comprehends, I still believe,

> […] all the knowhow, all the practices we learn just by growing up in a community, with almost everyone as a teacher, including social regularities like linguistic conventions, an especially important kind of social regularity. *This is huge and widely shared among the members of the community.* We do not agree or converge on meaning simply because we "grasp" somehow the same eternal "forms," but rather because we

> share a world in which we are integrated, and because we are worldly enough to see what to expect from each other. Our mutual linguistic understanding relies on that encyclopedic knowledge concerning how things are and how they are made, and on basic discriminatory capacities and abilities like mindreading and inferential ability. (Leclerc, 2009, 279).

For that reason, I do not believe that SLU correspond to a module of the mind. It does not satisfy two basic conditions for modularity: informational encapsulation, and domain specificity.

Now I want to suggest an idea that cannot be fully developed here. Consider for one moment the internal accusative of the verb "understanding". The list covers almost everything that is intelligible. We understand sentences, languages, cultures, books, face expressions, persons, attitudes, expectations, arithmetical series, problems, strategies, musical phrases, paintings, narratives and situations, physical systems, mechanisms, and certainly much more. My suggestion is that "linguistic understanding is only a part of it, and not an autonomous (modular) one. *Linguistic understanding relies on more primitive forms of understanding, specially the understanding of situations.*" (Leclerc, 2009, 274). It is not easy to say exactly what a situation is. A sentence-type describes a generic situation that may be found (or not) in the world. But we saw that the tokens of declarative sentences are always produced in a very specific situation, where the demonstrative conventions anchor the constituents of the sentence to specific, concrete objects, historical facts, etc. As a first approximation, I would say that a specific situation involves essentially agents with their expectations and plans, and *how things are and could be*. In ordinary language, an understanding of the specific situation (including the identification of expectations, intentions and plans through mindreading) is usually decisive for a correct understanding of the token produced in the context. I think Austin's theory of truth captures very well that difference, with its two types of conventions (descriptive and demonstrative). Take a sentence (token) from the specific situation where it has been produced, and many possible different understandings immediately come to mind. For instance, Putnam's example, "there is a lot of coffee on the table", can be true in very different settings, where "the way things are" might differ drastically (if there is a big coffee pot on the table, or if there are

bags of coffee beans on the table, or if someone spilt an entire coffee cup on the table, etc. We understand differently tokens of the same sentence because we have an understanding of the way things are in the specific situation we are in. A sentence taken from a specific context of use may serve different purposes from the one for which it has been uttered.

Now, consider the following sentences:

(a) You are very much Paris.
(b) Give peace a chance!
(c) Pride has a city!
(d) The Cardinal Mazarin has sent here his hemispheres.
(e) John speaks fluently English.

The first three are literal "nonsenses", but we do understand them easily. a) means that the addressee behaves very much like people do in Paris; b) means that we should try to create conditions for peace; c) that the people are proud of their city, and so are their servers in City Hall; d) is a malapropism. It really happened in Paris Parliament in the XVIIth century, but the deputy who uttered the sentence was immediately understood as saying: The Cardinal Mazarin has sent here his emissaries. e) is perceived as ungrammatical. We hear and understand easily a lot of ungrammatical sentences, especially from children. These cases are marginal, of course, but they reveal something interesting for my purpose. SLU is understanding of utterances, of actions performed for such and such a reason, and its first and primary object is occasion meaning. Occasion meaning is the result of a pre-propositional "modulation" (Recanati's expression) of standing meaning. The common use of language is a rule-governed rational activity, and rationality is precisely an unlimited capacity to realize intentional adjustments in a huge diversity of context. We revise our beliefs and plans constantly in the light of new information and evidence. This is what we do when we speak, and we speak in order to be understood. When we understand enough of the situation we are in, including the expectations and plan of the speaker, the words we hear are taken to carry the occasion meaning that most relevantly may contribute to the success of speaker's actions and plans at the moment of the utterance. It is that understand-

ing of the specific situation that enables us to understand easily and spontaneously ungrammatical sentences and malapropisms. If SLU were a kind of unconscious word processing mechanism, our ability to understand utterances involving deviant sentences would be seriously impaired.

André Leclerc
Federal University of Ceará-Fortaleza

References

Arnauld, Antoine & Nicole, Pierre. 1669-1672. *La Grande Perpétuité de la foi de l'Eglise catholique sur l'Eucharistie*, [*G.P.*], publiée par l'Abbé M***, Paris : Imprimerie de Migne, chez l'éditeur rue d'Amboise, Hors la barrière d'Enfer, 1841.

Austin, John L. Truth. In *Philosophical Papers*. Oxford: Clarendon Press, 1979.

Barber, Alex (ed.), *Epistemology of Language*, Oxford: O.U.P., 2003.

Barber, Alex. Truth Conditions and their Recognition. In *Epistemology of language*, 367-395.

Burge, Tyler. Comprehension and Interpretation. In *The Philosophy of Donald Davidson*, ed. by L.E. Hahn. Chicago and La Salle: Open Court, 1999.

Davidson, Donald. A Nice Derangement of Epitaphs. In *Truth and Interpretation: Perspective on the Philosophy of Donald Davidson*, ed. by E. Lepore. Oxford: Basil Blackwell, 1986. (Also in D. Davidson, *Truth, Language and History*. Oxford: Clarendon Press, 2005, 89-108.)

Davidson, Donald. The Social Aspects of Language. In *Truth, Language and History*, 109-128. First published in B. McGuinness & G. Oliveri (eds.) *The Philosophy of Michael Dummett*, Dordrecht, Kluwer Academic Press, 1994.

Davidson, Donald. 1967. Truth and Meaning. *Synthese* 17

Dummett, Michael. 1993. *Origins of Analytical Philosophy*. Cambridge (MA): Harvard University Press.

Evans, Gareth. Semantic Theory and Tacit Knowledge. In *Wittgenstein: To Follow a Rule*, ed. by S. Holtzmann & C. Leich. London: Routledge, 1981. (Also in G. Evans, *Collected Papers*. Oxford, Clarendon Press, 1985, 322-342.)

Grice, H. Paul. 1975. Logic and Conversation. In *Studies in the Way of Words*. Cambridge (MA): Harvard University Press, 1989.

Lakoff, George. 2004. *Don't Think of an Elephant!* White River Junction (Vermont): Chelsea Green Publishing.

Leclerc, André. Communication, Linguistic Understanding and Minimal Rationality in the Tradition of Universal Grammar. In *Logic, Thought and Action*, ed. by D. Vanderveken. Dordrecht: Springer, 2005, 133-150.

Leclerc, André. Meanings, Actions and Agreements. *Manuscrito*, Vol. 32, nº1, jan-jun, 2009, Special Issue: *Semantic Content and Communication*, ed. by A. Leclerc, E. Perini-Santos & M. Ruffino 249-288.

Longworth, Guy. A Plea for Understanding. In *New Waves in the Philosophy of Language*, ed. by S. Sawyer. Aldershot: Ashgate, 2009.

Millikan, Ruth. 2004. *Varieties of Meaning. The 2002 Jean Nicod Lectures*, Cambridge (MA): The MIT Press.

Millikan, Ruth. 2005. *Language: A Biological Model*, Oxford: O.U.P.

Moravcsik, Julius. 1998. *Meaning, Creativity and the Partial Inscrutability of the Hu-*

man Mind. Stanford: CSLI Publications.
Pettit, Dean. 2002. Why Knowledge is Unnecessary for Understanding Language. *Mind*, Vol. 111, n. 443, July 2002.
Polanyi, M. 1966/2009. *The Tacit Dimension*. Chicago: The University of Chicago Press.
Recanati, François. Does Linguistic Communication Rest on Inference? *Mind & Language*, Vol. 17, numbers1 and 2 February/April, 105–126, 2002.
Recanati, François. 2004. *Literal Meaning*, Cambridge, C.U.P.
Recanati, François. 2010. *Truth Conditional Pragmatics*, Oxford, O.U.P.
Schiffer, Stephen. The Mode of Presentation Problem. In *Propositional Attitudes. The Role of Content in Logic, Language and Mind*, ed. by C. Anthony Anderson & J. Owens. Stanford: CSLI Publications, 1990, 249-268.
Schiffer, Stephen. Knowledge of Meaning. In *Epistemology of Language*, ed. by A. Barber. Oxford: O.U.P., 2003, 303-324.
Soames, Scott. 2010. *What Is Meaning?* Princeton: Princeton University Press.
Travis, Charles. 2000. *Unshadowed Thought*. Cambridge (MA): Harvard University Press.
Travis, Charles. 2008. *Occasion-Sensitivity*, Oxford, O.U.P.
Wiggins, David. Meaning and Truth Conditions: From Frege's Grand Design to Davidson's. In *A Companion to the Philosophy of Language*, ed. by B. Hale and C. Wright. Oxford: Blackwell, 1997.
Wittgenstein, Ludwig. *The Big Typescript*. Oxford: Blackwell, 2005.
Wright, Crispin. Theories of Meaning and Speakers' Knowledge. In *Realism, Meaning and Truth*. Oxford: Blackwell, 1993, 204-238.

Relativo ma non troppo: de novo a favor da Objectividade na Execução Musical

António Lopes
Universidade de Lisboa e Grupo LanCog

BIBLID [0873-626X (2012) 34; pp. 739-753]

1 Introdução

Uma tese largamente difundida e aceite na comunidade musicológica é a de que os juízos de valor em música são fundamentalmente uma questão de gosto subjectivo moldado pelo contexto músico-cultural e pela tradição. Esta tese é principalmente reflectida nas análises de obras musicais, mas é facilmente transponível para o domínio da execução de tais obras, no qual ela consiste basicamente numa forma de pluralismo interpretativo, isto é, a ideia de que há várias interpretações incompatíveis de cada obra igualmente correctas e (esteticamente) boas, não havendo uma interpretação que seja a única maximamente correcta e boa.[1]

Esta situação contrasta com duas práticas bem estabelecidas na cultura musical actual no domínio da música clássica da tradição ocidental: a dos concursos para intérpretes e a da crítica de execuções, quer ao vivo, quer gravadas. Neste último domínio, em particular,

[1] Um exemplo: "I would like [...] to propose an indefinite moratorium on equations of analysis and value judgement, and indeed more broadly on equations of academic research and aesthetic approval. [...] It seems to me that the idea of the musical academy acting as some kind of quality control, with musicologists or theorists issuing admission tickets to a canonic hall of fame, is way past its sell-by date, and that the prerequisite for a more open-minded approach to musical culture than musicology has traditionally had is a more modest intellectual ambition: to register, to describe, to establish the facts as they are. [...] Of course, we cannot entirely eliminate our sense of what is good and bad from our professional work. But we can make a conscious attempt to moderate its influence" (Cook 2003: 261).

existe uma grande abundância de produção quotidiana de tipo mais ou menos jornalístico, evidentemente com níveis diferentes de competência e exigência. Como todas as actividades críticas, esta supõe a crença nalgum grau de objectividade nos juízos de valor estético, sem o qual não faria sentido que o crítico, para além de descrever uma execução, se pronunciasse quanto ao seu valor. De modo razoavelmente compreensível, são muitos os musicólogos que encaram com grande desconfiança uma actividade que não só é quase sempre pouco académica e científica, como ainda para mais adopta, implícita ou explicitamente, critérios de objectividade ou monismo estético. Tal desconfiança explica a recepção fria e talvez um pouco paternalista concedida por essa comunidade aos poucos estudos musicológicos com credenciais académicas que tentam demonstrar, mediante a análise musical, que algumas escolhas interpretativas são melhores do que outras na medida em que se adequam melhor à obra a executar.[2] Quanto aos concursos, eles são geralmente rejeitados, e não só por musicólogos, como eventos com um certo carácter circense, de interesse para a psicologia e a sociologia da música, nos quais pouco mais do que a habilidade técnica – e não o valor interpretativo - é apreciado e classificado.

Se afastarmos as razões de falta de competência, experiência, informação, sensibilidade ou parcialidade que colocam a maioria dos críticos de execuções de obras musicais num patamar demasiado abaixo do representado pela fasquia do crítico ideal de Hume, põe-se o problema: serão todas, ou pelo menos a grande maioria, das opções interpretativas, em última análise, igualmente defensáveis, sendo assim o seu valor estético ou artístico irremediavelmente relativo? Neste ensaio, e no seguimento de uma crítica anterior (Lopes 2006) às posições relativistas de Jerrold Levinson (Levinson 1987), defenderei uma resposta negativa, sugerindo que, à semelhança do que é defendido por muitos acontecer nas artes em geral,[3] há um

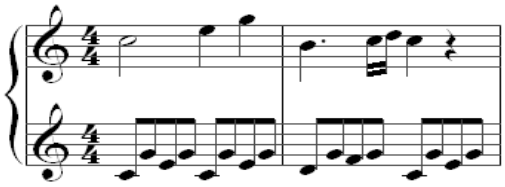
[2] Dois exemplos são Berry (1989) e Narmour (1988)

[3] Esta tese pode ser lida como pressupondo que se está a tratar execuções musicais (produto) como obras de arte. No entanto, é possível limitarmo-nos ao pressuposto mais fraco e pacífico de que elas (acto e/ou produto) são, pelo menos, arte – parte da arte musical. Para uma boa defesa de ambas as ideias, ver Kivy 1995, 122 ss.

espectro de objectividade ou universalidade na aceitação de juízos sobre o valor de execuções musicais, que vai de um extremo de objectividade (quase) total ao da relatividade (muito provavelmente) total. Para essa defesa serão apresentados alguns exemplos de domínios da execução musical nos quais certas estratégias ou escolhas interpretativas são claramente preferíveis a outras, sendo que, se essas o são, outras o poderão ser também. Os exemplos pretendem ser válidos pelo menos para a música tonal-modal, embora a sua transposição para algumas outras linguagens musicais da tradição ocidental seja relativamente linear. Isto bastará para demonstrar a falsidade da tese mencionada no início, e apoiar uma forma de monismo ou objectivismo *molto moderato* no campo da execução musical, alicerçado em princípios estéticos que, embora reconhecidos como suplantáveis ou fracos, no sentido de Sibley,[4] são, ainda assim, princípios.

1.1 Textura e condução de vozes: polifonia versus melodia acompanhada

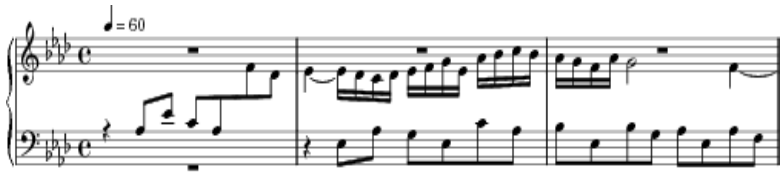
Uma distinção básica que abrange a maioria das obras da tradição ouvidas nas nossas salas de concerto diz respeito à textura da música. Por um lado, temos música que consiste numa voz principal, de carácter melódico, acompanhada por outras subordinadas, geralmente com ritmos e desenhos uniformes e relativamente simples, que as aglutinam entre si (usualmente em acordes, arpejados ou não) num plano de fundo e as distinguem da melodia. Um bom exemplo disso será uma secção de uma peça para piano do classicismo.



W.A. Mozart: Início da Sonata em Dó maior K. 545

[4] Sibley 1989

Por outro lado, temos as texturas polifónicas, em que as várias vozes gozam de grande independência nos planos rítmico e melódico, ainda que uma delas – por exemplo, a mais aguda – possa ser auditivamente seguida com maior facilidade. Para além disso, este tipo de textura articula-se muito frequentemente em função de temas (e contra-temas ou respostas), geralmente enunciados na entrada de cada voz, e do contraste entre, por um lado, momentos mais marcados pela presença desses temas etc., e, por outro, episódios mais, ou completamente, livres. Um bom exemplo disso será uma fuga barroca.



J. S. Bach: Fuga 17 do *Cravo Bem Temperado*, Livro I

Dadas estas duas texturas contrastantes, parece evidente que a escolha interpretativa que melhor as serve, e que é seguida por quase todos os intérpretes e educadores, consiste em dar maior saliência à voz que tem um papel mais destacado em cada uma dessas texturas. Assim, na melodia acompanhada, a melodia deverá prevalecer durante a maior parte do tempo, enquanto na fuga são as várias entradas das vozes que serão destacadas enquanto enunciam o tema, diminuindo de seguida a atenção a elas dada (o contratema beneficiará geralmente de uma saliência moderada, estando o tema simultaneamente presente noutra voz). Em geral, o destaque é gerido em função do material temático e da relação que a sua presença numa voz estabelece entre essa e as restantes vozes. Por exemplo, num *stretto*, em que as vozes entram com o tema começando na voz seguinte antes de a voz anterior ter acabado de enunciar o tema completo, sugerindo a imagem de um "atropelo" com acumulação de tensão, teremos de moderar a saliência de uma voz *antes* de ela terminar o tema. Por outro lado, secções da música em que não é apresentado o tema (nem o contratema) merecem menos destaque global.

É um facto que há regras específicas de cada contexto histórico-

musical que se aplicam à execução das obras nele criadas em termos de textura. No entanto, elas indicam-nos o *modo* de efectivar a estratégia interpretativa que é julgado adequado nesse contexto, e não que não deve haver distinção de relevância entre as vozes. Por exemplo, o modo mais óbvio de concretizar a hierarquia de vozes e/ou material temático é através da dinâmica: uma voz será ouvida como mais importante se soar mais forte do que as restantes. Este não é contudo o único meio, embora seja o menos subtil e mais eficaz. Ora, certos instrumentos, como o cravo e o órgão, não permitem tocar com maior volume as diferentes vozes que são tocadas no mesmo teclado. Contudo, a regra aplica-se aí também, com a diferença de que a saliência deverá ser realizada através da articulação (mais ligado ou mais destacado) e do *rubato* (pequenos atrasos ou acelerações). Por outro lado, o cravo, ou pelo menos alguns cravos, beneficiam também de uma diferenciação tímbrica de registos dentro de um mesmo teclado que, de um modo subtil mas real para o ouvido experiente, contribui para a hierarquização das vozes na ausência do recurso da dinâmica. Aliás, o modelo da forma polifónica é, em grande parte do repertório, e certamente no anterior ao séc. XIX, a música vocal, e nesta é sempre possível, por recursos dinâmicos, agógicos, tímbricos e outros, conferir destaque à voz ou vozes desejadas.

Do mesmo modo, também no caso da textura de melodia acompanhada há pelo menos uma escolha interpretativa – dar ou não relevo à melodia (principal) – que quase não é escolha alguma, na medida em que não o fazer seria visto como falta grave de musicalidade, ou mesmo incompreensão da natureza da obra musical ou do excerto em causa. Este carácter objectivo e artisticamente imperioso da atenção à melodia na textura em questão reforça pelo menos um monismo interpretativo moderado, sendo contudo certamente compatível com o relativismo em outros aspectos relacionados. Assim, a importância da melodia tem de ser evidente, ainda que haja momentos em que faz sentido suavizá-la um pouco em proveito de algo musicalmente relevante a ocorrer simultaneamente numa melodia secundária ou mesmo na estrutura harmónica ou rítmica do acompanhamento. Além disso, diferentes intérpretes valorizarão de diferentes maneiras as melodias secundárias eventualmente formadas no seio do acompanhamento, de entre as quais devemos ainda distinguir as que mais claramente foram intencionadas como melodias de algum relevo pelo

compositor e aquelas que o próprio intérprete realmente descobre na textura. Finalmente, aqui como no caso da textura polifónica, o repertório de meios musicais que é possível empregar na diferenciação e condução de vozes está mais ou menos objectivamente delimitado, mas as escolhas dentro desses limites fazem parte do domínio da liberdade artística do intérprete, com resultados por vezes mais, por vezes menos, difíceis de antever num raciocínio indutivo, i.e., mais ou menos regrados por princípios universalizáveis, tal como sucede com a escolha dos meios usados para realçar os temas ou para assinalar a diferença entre exposição e episódio numa fuga.

Repare-se que este tipo de escolha interpretativa está alicerçado em propriedades do conteúdo melódico e rítmico da própria música – incluindo as características gerais que fazem de algo uma (boa) melodia tonal ou modal – e em realidades da nossa percepção auditiva – não é possível ter sempre a mesma atenção auditiva a todas as vozes, ou pelo menos, não é desejável interpretá-las desse modo –, e não em padrões de gosto pessoal ou contextual. Todo o ouvinte competente deve ouvir como monótona, para além de confusa, uma execução que apresente a obra como desprovida de qualquer destaque e diversidade nos parâmetros referidos.[5] Podemos então falar, parece, de um princípio consideravelmente geral que poderíamos baptizar como "o que é interessante deve ser relevado", de onde se derivam princípios mais específicos como o relativo à execução de música com as texturas que acabamos de analisar. A projecção da sinergia entre informação e diferenciação conseguida mediante a aplicação destes princípios mais específicos poderia ser interpretada, em termos dos cânones de Beardsley, como favorecendo um equilíbrio entre variedade e unidade que resultará numa maior intensidade expressiva. Assim, não faria sentido alguém usar este tipo de diferenciação, quando aplicado correctamente, i.e., às vozes correctas nos momentos correctos, como razão para justificar um juízo estético negativo sobre a execução em causa.

[5] Tal opção pode ser uma virtude em casos como o de alguma música minimal repetitiva, e *talvez* em *algumas* obras neo-clássicas, mas trata-se evidentemente de excepções.

1.2 Clarificação da textura orquestral

Outro princípio específico relativamente ao qual parece haver uma larga concordância entre críticos, senão mesmo unanimidade, entre músicos e críticos, é a de que uma execução de música de conjunto, com especial relevo para a orquestral, que consegue clarificar em grau elevado a textura, os timbres, as vozes polifónicas e suas gradações dinâmicas, etc., é, *ceteris paribus*, uma melhor execução do que uma que não o consegue ou o consegue em grau inferior.

Para avaliarmos esta tese, convém situarmo-nos num segmento do repertório em que poderia parecer que tais qualidades são, não uma virtude, mas um defeito da execução. Tanto quanto me é dado ver, o melhor candidato seria, por exemplo, uma obra orquestral grandiosa, sombria e solene do romantismo tardio.[6] Quando é que a clarificação das texturas seria uma falha na sua execução? Apenas, parece-me, se essa clarificação for inevitavelmente prejudicial a valores importantes para o estilo em causa (por exemplo, a sensação de peso orquestral, especialmente nos graves, a grandiosidade e a acumulação/resolução de tensão nos clímaxes, o carácter vibrante e o *vibrato* propriamente dito) ***e*** não trouxer outros atributos ao nível dos elementos da textura que evidencia ao clarificar, que possam contrabalançar as perdas nos elementos importantes acima referidos. Assim sendo, esta excepção é de certo modo também ela dotada de regras, não estando completamente mergulhada em relativismo. De qualquer modo, em todos os outros casos, que são a grande maioria, parece absurdo afirmar que uma dada execução é pior pelo facto de clarificar a textura polifónica de uma obra – mesmo a de uma do repertório acima mencionado, sendo a dificuldade em obter nela tal clareza muitas vezes um factor de desafio aos intérpretes que lhes granjeia mérito adicional da parte da crítica, no caso de serem bem

[6] Para quem não possa passar sem um exemplo concreto, talvez a 8ª Sinfonia de Bruckner, ou a maior parte das passagens orquestrais do *Crepúsculo dos Deuses,* de Wagner. A sugestão da música orquestral do impressionismo francês poderia parecer ser também um terreno adequado para a colheita de contra-exemplos. No entanto, ele é menos propício do que é aparente, dada a preferência dos compositores franceses (especialmente Ravel) e dos seus intérpretes nativos por texturas transparentes e pouco pastosas, e pela moderação nos *tempi*, como noutros parâmetros.

sucedidos. Afirma-se, isso sim, que uma tal execução é má porque, à força de querer clarificar texturas, perde em características do tipo das acima referidas no repertório adequado e quando estas têm um peso elevado.[7] Se tudo permanecer igual, porém, uma execução só poderá, em geral, ser melhor por obter essa clarificação, expondo à atenção do público mais detalhes composicionais interessantes que de outra forma permaneceriam obscurecidos. Com efeito, é até defensável que mesmo nos casos em que a clarificação revele alguma superficialidade ou falta de sofisticação na textura de uma obra, a revelação seja ainda um mérito enquanto qualidade da execução, por nos permitir analisar e avaliar com mais informação e maior correcção e justiça a obra executada.[8]

Um bom exemplo da quase universalidade deste tipo de preferência pode ser encontrado no sucesso do chamado movimento da "interpretação historicamente informada" (ou, no seu infeliz título original, felizmente evitado hoje em dia, movimento da "interpretação autêntica"). A procura dos meios e estilos de execução que repliquem as condições originais dos contextos de criação e estreia das obras teve certamente a seu favor muitos componentes de carácter ideológico.[9] No entanto, com o passar do tempo e a perspectiva histórica sobre o próprio movimento, podemos afirmar com segurança que as marcas características que ele nos trouxe de novo face ao paradigma tardo-romântico – entre outras, texturas mais transparentes, articulação mais pronunciada, maior relevo de tempos fortes no compasso, grande atenção ao texto cantado – não só foram acolhidas por praticamente todos os intérpretes desse reportório e pelo público, passando a ser a norma, como inclusivamente transbordaram as fronteiras da música pré-romântica, tornando-se, algumas delas, exigências genéricas, ainda que naturalmente com maior incidência em certos repertórios do que noutros.

---

[7] Uma possível excepção – mas apenas possível – será a de obras como algumas de Ligeti ou Penderecki num estilo geralmente apelidado de música de texturas.

[8] Para um desenvolvimento e defesa da ideia de que há propriedades de valor positivo nas execuções de obras que não derivam directamente de propriedades, positivas ou não, das obras executadas, ver Lopes 2005.

[9] Desenvolvo este e muitos outros aspectos do movimento em Lopes 2010, cap. 10.

Não parece tratar-se somente de uma questão de gosto actual, uma vez que mesmo o público melómano (provavelmente) minoritário hoje em dia que aprecia as suas execuções das obras relevantes do romantismo tardio servidas com uma dose extra de peso orquestral, profundidade expressiva, timbres escuros e carácter sombrio, e o não tão minoritário grupo a quem desagrada a excessiva leveza e brancura de algumas vozes agudas típicas dos primeiros passos do movimento em causa, não deseja ver sacrificados a esses valores mais subjectivos os méritos mais universais, na dose certa, que o sucesso do movimento trouxe, por arrastamento, ao repertório em causa.[10] Na verdade, as vozes críticas em relação a alguns resultados do movimento, que não as motivadas apenas pela falta de hábito, erguem-se quase sempre contra o exagero nalgumas características e opções do mesmo que põem em causa o equilíbrio relativamente a outros valores também eles consensuais.[11] Mesmo a excepção mais evidente nesse sucesso dos instrumentos e práticas contextuais, a resistência à substituição do piano pelo *fortepiano* no repertório que vai do clas-

[10] Peço a compreensão do leitor para esta aparentemente abusiva assunção da legitimidade para falar em nome dessa(s) minoria(s) com a confissão de que se houvesse uma estrutura institucional para tal sensibilidade em matéria de execução musical, eu seria provavelmente um dos seus membros fundadores.

[11] A este título, refira-se como exemplo a nostalgia bastante compreensível pelas execuções da música sacra de Bach por intérpretes da antiga *mainstream* como Karl Richter, patente por exemplo nas páginas da revista de referência *Gramophone* (em especial, as assinadas pelo crítico da especialidade, Jonathan Freeman-Attwood). Essa crítica às execuções "historicamente informadas" não se daria se não fossem as falhas de várias execuções desse tipo dos nossos tempos (mas sobretudo dos anos 80 e princípio de 90) em termos de expressividade na enunciação dos textos, que parece sacrificada à uniformidade e à segurança técnica mais ou menos asséptica (para além de algum sentimento de banalização da arte, patente na ideologia do movimento inicial, que se pode materializar de modo infeliz nessas características, e em casos como a admissão ao estatuto de execução profissional de intérpretes que não estão à altura das exigências, em virtude da "especialização" ou de atractivos até certo ponto extra-musicais (caso de muitos contra-tenores substituindo meio-sopranos e contraltos bem mais seguros e competentes, graças à aura de historicidade da técnica do falsetista – cf. nota 9). Ou seja, não tem de haver incompatibilidade de princípio entre os valores da execução musical estabelecidos, como os da musicalidade e da expressividade, e a clareza de texturas e articulação, etc., ainda que a harmonização entre essas duas "classes" de valores seja na prática difícil de conseguir (se é que o é de facto).

sicismo até meados do séc. XIX, parece mostrar uma certa consistência e selectividade saudável no acolhimento das "novidades" da "interpretação historicamente informada", e até ser sintomática do que tenho vindo a expor. A menor aceitação pública do instrumento como alternativa ao piano moderno (não estou a falar do meio da musicologia histórica, naturalmente) mostra que não é desejável a clareza e leveza a qualquer preço, e que, para além de um som pouco consensual da parte do instrumento antigo, o facto é que a sua contraparte moderna consegue, quando apropriadamente tocado, proporcionar praticamente toda a clareza e contraste exigidos pelo repertório do *fortepiano*, com evidentes ganhos tímbricos, mecânicos e de paleta dinâmica. Assim, o saldo desta manutenção da preferência pelo piano é positivo, não indo no entanto realmente contra a desiderabilidade da maior clareza de texturas.

2 Projecção da estrutura formal

Uma conclusão a extrair da leitura da literatura crítica de melhor qualidade é a de que uma eficaz gestão das *propriedades formais ou estruturais* da obra, i.e., a capacidade de um intérprete para projectar na audição do público os elementos da forma da obra e as suas articulações, é um dos principais méritos da execução de obras musicais. Um dos melhores exemplos da aplicação deste critério avaliativo é o domínio da gestão dos diferentes patamares de tensão e resolução, de modo a que o clímax principal de uma obra ou secção não seja nivelado com clímaxes secundários. Outro exemplo é o da chamada de atenção por parte do intérprete, na sua execução – e por meios diversos (dinâmica, *rubato*, etc.) –, para momentos centrais na articulação de formas musicais, como a recapitulação numa forma-sonata ou a transição de uma variação para a seguinte.

Naturalmente, não é o caso que, em todas as obras de todos os géneros da história da música clássica ocidental, as componentes estruturais sejam igualmente relevantes, nem do ponto de vista da pura análise composicional, nem, sobretudo, da perspectiva das possibilidades (e do interesse artístico) da sua realização em execuções. Assim, será mais provável que um crítico apele a critérios desta categoria na apreciação de uma execução de uma sinfonia de Bruckner do que na de um prelúdio de Chopin. Por outro lado, embora a com-

ponente formal seja comparativamente tão importante na sinfonia como num moteto isorrítmico de Dufay, as limitações à projecção em execução no caso deste último são de uma natureza muito diferente, razão pela qual o apelo crítico a tal projecção – se ele existe de todo – é também bastante diferente. Como seria de esperar, expressão e forma não são completamente dissociáveis, havendo relações evidentes entre o modo como ambas se manifestam em execuções. Para ver isso, basta considerar o facto de que um dos recursos mais fundamentais de que o intérprete dispõe para delinear e projectar para um público a estrutura de uma obra ou secção consiste precisamente no doseamento da expressividade a aplicar a cada parâmetro, elemento ou evento musical.

Também aqui, parece-me, estamos na presença de um princípio – projectar a estrutura formal de uma obra – que só pode funcionar a favor do mérito de uma execução, ou, pelo menos, *quase sempre* funciona a favor desse mérito. Se duas execuções diferirem, *grosso modo*, "apenas" no facto de que uma projecta eficazmente essa estrutura e a outra não, será altamente provável que a que o faz será a melhor das duas. Trata-se, neste caso, e até de um modo mais evidente do que nos princípios anteriormente analisados, de um princípio que abona a favor do valor positivo da execução *mesmo* quando a estrutura formal da obra assim revelada não é dotada de grande interesse, sofisticação ou originalidade. Com efeito, pelo menos nas formas mais tradicionais – sonata, rondó, canção, variação, etc. -, a forma deve ser do conhecimento dos ouvintes experientes antes da escutarem uma nova obra, tratando-se antes de comparar o modo como um intérprete em particular projecta essa forma com o modo como outros o fazem, ou com o falhanço nessa projecção. Mais do que a aplicação de uma forma nova (ainda que isso possa ser o caso), o que importa do ponto de vista da forma musical estabelecida é compreender e apreciar o modo como o compositor tece o seu discurso musical em torno de um esqueleto pré-determinado. Mostrar isso ao público de um modo compatível com a musicalidade e a criatividade – ou mesmo potenciado por estas qualidades – é uma virtude da execução musical.[12]

[12] Não quero aqui ignorar o facto de que há discordância entre os musicólogos acerca da interpretação correcta da forma de algumas obras musicais do repertório, e logo, que ela existe entre os intérpretes dessas obras, pelo que o pluralismo

Se precisarmos de um princípio mais geral de onde este possa ser derivado, à semelhança do que fiz acima, poderíamos sugerir que a virtude artística servida pelo meio da projecção da estrutura formal é em geral o de que algo que guia correctamente o ouvinte na sua audição é sempre bem-vindo. Claro que, como muitos outros princípios, e talvez todos em estética, este conhece excepções e pode ser suplantado por outros princípios quando não houver plena compatibilidade entre todos eles. Assim, uma execução em que o intérprete se esforce de tal modo nessa projecção que ponha em causa detalhes e efeitos importantes – por exemplo, à força de querer guiar o ouvinte para a identificação dos clímaxes mediante a subordinação de tudo o resto – poderá, em função do repertório, estilo, e outras considerações, ser uma execução menos boa do que uma em que a atenção ao detalhe acompanhe uma atenção não tão grande à estrutura. No entanto, como já referi, estas excepções só seriam perigosas para uma posição mais fortemente objectivista do que a moderada que aqui defendo.

3 A dicção na música vocal com texto

Um último e breve exemplo de propriedade de execuções cuja posse só pode ser usada como razão a favor do valor positivo da mesma é o da clareza na dicção do texto cantado. Seja em música para solistas ou para conjunto, serão extremamente raros os casos em que essa propriedade possa ser usada para fundamentar um juízo negativo acerca da execução. Na verdade, acontece por vezes que essa qualidade é incompatível com outras qualidades igualmente positivas do canto, como a beleza do timbre, a largueza do espectro de dinâmicas usadas, a consistência e uniformidade da voz, e, mais particularmente, com o *legato* próprio de todo o canto tecnicamente educado na tradição em foco. Contudo, o facto contingente de que para alguns cantores é difícil, ou mesmo impossível, obter clareza na dicção do texto sem resultados pesadamente negativos no *legato* ou na emissão de notas agudas não deve ser usado como um argumento contra o

interpretativo será o caso nestas situações (agradeço ao Prof. John Rink por me alertar para este ponto (em comentário pessoal)). Tudo o que necessito de afirmar é que, na maioria dos casos, a compreensão da forma é relativamente pacífica, i.e., interpretativamente monística.

princípio estético geral de que uma execução vocal é melhor se, em adição aos méritos que já exibe, também o texto for correctamente enunciado e perceptível pelo público, e nunca é pior por isso. Tal dificuldade poderá, isso sim, ser um argumento a favor da necessidade de aperfeiçoamento técnico da parte dos músicos em causa. Este princípio pode assim ser contado entre os princípios "fracos" ou suplantáveis conducentes ao maior valor estético ou artístico da execução musical, e talvez mesmo entre os menos fracos e suplantáveis de entre aqueles.[13]

4 Uma objecção respondida: circularidade?

Em conclusão, resta-nos considerar uma objecção pertinente à defesa do monismo moderado que apresentei. Consiste ela em detectar, nas expressões usadas para denotar as propriedades aqui referidas como sustentando esse monismo, um elemento de petição de princípio. Segundo a objecção, tais expressões contêm já aspectos valorativos de polaridade positiva. Assim, quando aqui se defende que as propriedades (razoavelmente) objectivas ou puramente descritivas denotadas por expressões tais como "clareza da polifonia/textura orquestral", "saliência dinâmica das entradas do tema da fuga" ou "relevo expressivo da melodia (acompanhada)" só (ou quase só, *ceteris paribus* e garantido o repertório adequado) poderão contar como valorizando positivamente as execuções que exibem tais propriedades, o oponente diria que a conclusão já estava de algum modo contida nas premissas, uma vez que, supostamente, só em virtude de uma avaliação positiva prévia se pode atribuir com correcção essas propriedades a execuções.

Contudo, não me parece que a objecção colha. Com efeito, qual das expressões da pequena lista acima contém um elemento indelevelmente avaliativo? Clareza? Mas o predicado 'claro' atribuído, por exemplo, a uma voz (cantada), é puramente descritivo. Que Emma Kirkby tem uma voz clara e Jessye Norman tem uma voz escura são afirmações (metaforicamente, talvez, mas contextualmente con-

[13] Observe-se que, sintomaticamente, a enunciação clara e perceptível do texto parece ser reconhecida pelos críticos como uma virtude performativa mesmo quando esses críticos, ou o público, não dominam a língua do texto a ser cantado.

sensualmente) verdadeiras que não comprometem ninguém que as afirme com juízos de valor – embora, dependendo das sensibilidades individuais, possam justificar que será provável que certos sujeitos apreciem mais uma dessas vozes de soprano do que a outra. Mais a propósito, 'claro' aplicado a uma textura musical parece ser tão puramente descritivo quanto a aplicação acima, sendo um facto contingente que (praticamente) todos os ouvintes competentes apreciem uma maior clareza na textura (com as qualificações já referidas). Tal facto pode mudar no futuro, mas o meu objectivo não é evidentemente defender que é uma verdade conceptual ou analítica que clarificar de uma textura em execução é melhor do que não o fazer, nem sequer que se trate de algum universal estético – embora isso não seja propriamente impossível. O mesmo é o caso relativamente a propriedades como as referidas pelas expressões 'saliência' – é algo *ipso facto* esteticamente bom só por ser saliente? – e 'relevo expressivo' – nem todo o parâmetro musical afecta positivamente o valor de uma execução apenas porque lhe é dado relevo expressivo, e alguns contribuem mesmo negativamente se lhes for conferido tal relevo.[14]

A objecção poderá, não obstante, ter maior força se incidir sobre o princípio mais geral que enunciei como uma fonte de onde poderiam derivar pelo menos alguns dos princípios mais particulares acima, o princípio que afirma que o que é musicalmente interessante na composição deve ser evidenciado em execução. O predicado 'interessante' tem evidentemente uma carga valorativa. No entanto, creio que é possível mostrar que há critérios razoavelmente objectivos e universais para identificar, dados os parâmetros relevantes, que eventos ou propriedades musicais são interessantes. De facto, trata-se de algo que é aprendido desde cedo na educação musical, quer do ouvinte, quer do intérprete. A franja de relatividade na determinação do que é musicalmente interessante é real, mas não demasiado larga. Critérios objectivos, ou pelo menos largamente consensuais, para identificar graus de relevância musical de elementos como a presença de um *cantus firmus*, a existência de linhas melódicas dignas de atenção no interior de uma textura ou a antecipação de um tema de uma secção climática numa secção anterior estão em geral disponíveis a

[14] Suponha-se, por exemplo, que um intérprete decide dar relevo dinâmico aos ritmos pontuados num *Nocturno* de Chopin.

uma boa parte do público musicalmente educado e experiente em mais do que uma audição informada de cada obra, sem o prejuízo de que abordagens mais sofisticadas – musicológicas – possam revelar algumas qualificações a tal objectividade, mas apenas em alguns casos, não certamente na maioria.

António Lopes
Universidade de Lisboa e Grupo LanCog

Referências

Berry, Wallace. 1989. *Musical Structure and Performance*. New Haven: Yale University Press.

Cook, Nicholas. 2003. Writing on Music or Axes to Grind: road rage and musical community. *Music Education Research*, Vol. 5, No. 3, Novembro 2003, 249-261.

Kivy, Peter. 1995. *Authenticities. Philosophical reflections on musical performance*. Ithaca NY: Cornell University Press.

Levinson, Jerrold. 1987. Evaluating Musical Performance. In Jerrold Levinson, *Music, Art and Metaphysics: Essays in Philosophical Aesthetics*. Ithaca NY: Cornell University Press, 376-392.

Lopes, António. 2005. Musical Works and Performance Evaluation. Postgraduate Online Journal of Aesthetics (British Society of Aesthetics), 2: 2 (Agosto 2005)
(https://leopard.ncl.ac.uk/BSA/PJAArchive.aspx)

Lopes, António. 2006. Relativismo na Avaliação de Execuções Musicais. *Philosophica* (Revista do Departamento de Filosofia da Universidade de Lisboa), nº 27.

Lopes, António. 2010. *O Valor de um Bach Autêntico; um Estudo sobre o Conceito de Autenticidade na Execução de Obras Musicais.* Lisboa: Fundação Calouste Gulbenkian.

Narmour, Eugene. 1988. On the Relationship of Analytical Theory to Performance and Interpretation, in *Explorations in Music, the Arts, and Ideas*, ed. by E. Narmour and R. Solie. New York: Pendragon Press, 1988, 317-340.

Sibley, Frank. 1989. Aesthetic Concepts. In *Aesthetics. A Critical Anthology*, ed. by George Dickie et al. New York: St. Martin's Press, 356-374.

Wittgenstein on Mathematical Identities

André Porto
UFG – Brazil

BIBLID [0873-626X (2012) 34; pp. 755-805]

I - Introduction

This paper offers a new interpretation for Wittgenstein`s treatment of *mathematical identities*. As it is widely known, Wittgenstein`s mature philosophy of mathematics includes a general rejection of *abstract objects*. On the other hand, the traditional interpretation of mathematical identities involves precisely the idea of a single abstract object – usually a number –named by both sides of an equation. We may follow (QUINE, 1960, p. 114) and take identity as a relation directly between terms or we may prefer Frege's ideas and insist on its being mediated by senses (FREGE, 1977, p. 56). Either way, we normally take these statements as crucially involving an appeal to a *single object* as the referent of both *relata*, weather they are senses or terms. But if we completely let go the idea of this abstract pivot, how could we possibly make sense of such mathematical statements? What would sentences such as "$2 + 3 = 5$" *be about*, if they were not about *numbers*?

This is where the idea of *back and forth correction* comes in. We propose this terminology to refer to Wittgenstein's notorious revamping of the traditional construal of (mathematical) identities. As we shall see, this new interpretation of equations also touches upon at least two other central elements of his philosophy of mathematics. The first element is his treatment of mathematical *generality* (and his rejection of *quantifiers*[1]). The second one is his handling

1 It is perhaps important to remember here that Wittgenstein is not alone in this rejection of quantifiers. Skolem (SKÖLEM, 1952,, p. 120) also singles them

of statements involving *Infinity*. Taken together, his treatment of mathematical identities, of generality and infinity, make up a large portion of Wittgenstein's approach to mathematics, a proposal which is distinct both from the classical realistic proposal and from the constructive one, centered on the notion of potentiality.

Most features of this alternative position regarding mathematics are still not widely known and thus his whole proposal remains unacknowledged by researchers on the foundations of mathematics at large.[2] We believe this sad situation is in part due to the difficulty of dissociating Wittgenstein's ideas from his very idiosyncratic jargon, thus making them available to a wider audience, and in part due to our sheer lack of understanding of the key logical elements of that approach.

Our article will be divided into three parts. We will begin by a general discussion of some other very general approaches to mathematical identities. Our goal will be a negative one. Throughout our paper we will use these other alternative handlings of identities in order to contrast them with Wittgenstein's proposals. In the second part of our paper we will focus on Wittgenstein`s extensive analysis of a special type of mathematical identities, the ones involving *recurring decimals*. Two main ideas will be introduced here. The first one will be his critique of the *elision dot`s notation*. The second will be his insistence on the introduction of the syntactical operation of *iterated copying* on the right side of that kind of equation. We will dedicate the entire final section of our article to Wittgenstein`s treatment of mathematical identities as *back and forth correction*. This will involve a generalization of the treatment of recurring decimal's identities to other kinds of mathematical identities.

Before we move on to the main argument of the paper, a note on exegetical claims. Wittgenstein's writing is very extended and varied. There is no agreement on just what he said (or implied), and when. Many times we don't even know if he is the one doing the

out as the key source of the fundamental "faults" afflicting classical mathematics.

2 Despite important monographs on the theme, such as (WRIGHT, 1980), (SHANKER, 1987), (STENLUND, 1990), (FRASCOLLA, 1994) and (MARION, 2008).

talking, or his ever present interlocutor. Any bolder reconstruction such as the present one runs the risk of appearing to be more than just that: plain misattribution to him of views he never sustained. As we said before, our main objective is to try to view his philosophy from *without*, offering as wide a perspective as possible of his work within the foundations of mathematics. To be able to do so with a little more freedom, we will follow an option that has become common ever since Kripke wrote (KRIPKE, 1972).[3] We will drop any substantial exegetical claims. We still try to offer (what we take to be) the appropriate quotes, but the reader is free to disregard them if he or she thinks they are not nearly enough to sustain our exegetical claims.

The Operational Reading of Mathematical Identities

Let us take a simple mathematical identity such as

(1) $$2 + 3 = 5$$

If we lay aside all our mathematical and logical sophistications aside and, say, go back to an elementary school reading of such statements, it appears natural to take these identities as being *directed*. We *start* from the left, by operating with the numbers 2 and 3 – we add them – and then we *obtain*, or produce, the result, in the right hand side of the equation, the number 5. In such a reading, the two sides of the equation are *not on the same level*. The interpretation is not "*flat*". It has a (main) *orientation*: we start from the *left* side and *generate* the result on the *right*.

$$2 + 3 \rightarrow 5$$

This is the reading that is preserved, of course, in a *recursive* interpretation of (1). By utilizing the well known definition of addition

$$\begin{cases} a + 0 = a \\ a + sn = s(a + n) \end{cases}$$

3 Cf. also (STENLUND, 1990, p. viii)

we could even represent this operation as a sort of "short hand" for series of step by step "externalizations" of the unary operation of "successor for":

$$ss0 + sss0$$
$$s(ss0 + ss0)$$
$$ss(ss0 + s0)$$
$$sss(ss0 + 0)$$
$$sssss0$$

This way of viewing identities is precisely the one advocated by intuitionists such as Martin-Löf (MARTIN-LÖF, 1984, p. 71). In such readings, there is an *intrinsical unbalance* between the result, $sssss0$, and the initial left hand side expression, $ss0 + sss0$. This latter expression is not in *canonical notation*. The very idea of an "*operation*" is explained by appealing to the fundamental distinction of *canonicity* and *non-canonicity* (DUMMETT, 1991, p. 175; MARTIN-LÖF, 1987; PRAWITZ, 1977). We could even say that $ss0 + sss0$ *denotes an number*, but only because this *non-canonical* expression *could be transformed* into a *canonical* expression, the expression $sssss0$. This is the semantical intuitionist's view of the notion of "number". [4]

Before we leave behind the constructive approach to arithmetical identities and move on to Frege's ideas concerning such statements, let us quickly draw our attention to one important ingredient of the constructive approach to mathematics, an aspect that laid somehow hidden in our seemly innocuous modal expression "*could be transformed*" (in the last sentence of our previous paragraph). The duty a semantical intuitionist expects this modal to perform is anything but trivial and innocuous, philosophically speaking. The problem is: any non canonical expression can *succeed in denoting* a *number* only by appealing to its "*possible transformation*" into a canonical expression of the form "$s \ldots s0$". It's true that one can *prove*,

4 We will only consider in this paper the contemporary type of intuitionism of Dummett, Prawitz and Martin-Löf, usually known as "Semantical Intuitionism". We will not discuss the classical intuitionism of, say, Brouwer.

within an intuitionistic system such as Martin-Löf's, that these operations are "possible".[5] But, for numerical expressions involving "very large numbers" (say, as $10^{10^{10}}$), how possible is this posited possibility?

The idea of having the very *semantics of mathematics* depend on these grossly idealized possibilities (the "impossible possibilities") appears suddenly as philosophically fishy. One should not forget here also that the central argument for the intuitionistic assault on classical mathematics had to do precisely with criticisms regarding idealized possibilities such as these:

> The fact that quantification over an infinite totality shares so much in common with quantification over a finite one tempts us to overlook the crucial difference from the case in which we are quantifying over a finite totality *which we can survey*, namely that we do not have, even in principle a method of determining the truth-value of quantified statement by carrying out a *complete inspection of the elements of the domain and checking for each one, whether the predicate applies* (DUMMETT, 1977, p. 6).[6]

Thus it should not come as a surprise that authors, such as Crispin Wright, should pointed out that "arguments essentially analogous to those which the mathematical Intuitionists … use to support their revisions of classical logic" could now be redirected against the intuitionists themselves (WRIGHT, 1993, p. 107).[7] We will come back to this point latter on.

Frege's "flat" reading of identities

For a platonist such as Frege, a number is a number, not because we can prove it to be so (within some formal system), but because

5 This is precisely the task of N-elimination rules in Martin-Löf's system (MARTIN-LÖF, 1984, p. 71)

6 My italics.

7 For more extensive discussions of this point, cf. (MARION, 2008, p. Chap 8; STENLUND, 1990, pp. 146-51)

mathematical reality is such. Numbers are abstract objects whose existence is conceptualized as being completely independent of any epistemological act such as a demonstration. Accordingly, in this classical reading of identities anticipated in the beginning of our paper, $2 + 3$ is equal to 5 simply because both "$2 + 3$" and "5" *name the same object*. This naming relation is not tributary to any more fundamental *operation,* such as the successor function.

As we know, Frege (and Russell) famously advocated a purely *cardinal view* of numbers as equivalence-classes generated by the equinumero-sity-relation (FREGE, 1964, p. 78). In sharp contrast to the *ordinal* approach latter favored by the intuitionists, any number such as the number 3 is not seen as the result of, say, there applications of the operation successor to an initial object 0

$$s\Big(s\big(s(0)\big)\Big)$$

but as an enormous class of all "triples".

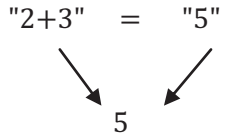
Thus, in contrast to the previous, constructive alternative we presented before, in the classical, fregean view, a mathematical identity is something completely "symmetrical". There is no prior notion of a "result" (of an operation). "5" names an abstract object in precisely the same way as "$2 + 3$". And since both of them happen to name the same abstract object, the statement (1) is true. 5 is "as much equal" to "$2 + 3$" as vice-versa, there is not a more fundamental "direction" here.[8]

Following (BAKER & HACKER, 1988, p. 22) I propose to call this classical view of identities "*flat*". This is important because, as we will see, contrary to what one might have expected, even thought Wittgenstein does insist on introducing the notion of "operation", his reading of identities is *completely flat*. More than that, as Backer and Hacker correctly pointed out, this "flat view" is characteristic not only of his interpretation of identities but of his

8 We could of course insist that the name "5" is *simple* and that "$2 + 3$" is a *complex name*, but this would appeared to be system-dependent. We would still need some sort of "transcendent" or "metaphysical" notion of *simplicity* here to mark out the difference.

whole approach to mathematics, all the way from the *Tractatus* on (WITTGENSTEIN, 1961, p. 6.127). This is but one of the fundamental differences between Wittgenstein and the constructivists.[9]

One has to be careful here, thought. Wittgenstein reading of (1) is *flat* in the sense of its *not being directed*. But that does not mean that Wittgenstein endorses the *classical* "*triangular*" view of mathematical identities (as true when both its side-names denote the same abstract object).



He never did accept the triangular construal. There are two notions of *identity* in the *Tractatus*. One of them is the "*ontological*", it is asserted *directly of objects*, not of *names* or *senses* of objects. But this notion degenerates into mere self-identity, $[\lambda x.x = x]$ (WITTGENSTEIN, p. 5.5303). The other one is mere *synonymity* (WITTGENSTEIN, p. 6.2323). Neither of them, though, is the classical, "triangular" notion. In fact, as we've anticipated before, Wittgenstein all together *rejected* any idea of an "abstract object" (acting as *pivot* for identities). As we pointed out, for him mathematical statements *are not* about any such "abstract entities", such as *numbers*. And so, *per force*, his rendering of mathematical identities *could not possibly be* the classical one.

II - Recurring Decimals
Criticism of the elision dots' notation

Following Wittgenstein, let us change our main example from the simple equation "$2 + 3 = 5$" to something a little bit more complicated: identities involving infinite recurring decimals such as:

(2) $$1 \div 7 = 0.142\ldots$$

9 Another important difference is his rejection of any notion of "in principle possibility" and accordingly of any idea of "potential infinity".

The reason for Wittgenstein focusing on these (rather plain) mathematical identities is quite apparent. In (2) we have pretty much the first (and simplest) example of *statement involving infinity* in our elementary mathematical education. In a typical strategy for him, the philosopher finds here a thoroughly *perspicuous* example on which to focus his investigation. In examples such as this, he has all the key elements of some general problem he is interested in without any inessential technical complexities. As we will see, in the case of recurring decimals, these ingredients include *mathematical identities and abstract objects*, *generality* and of course, *infinity*.[10]

The first element Wittgenstein is interested in his surprisingly long and elaborate discussions on *recurring decimals*[11] is of course the *elision dots' notation*. This is a theme that absorbed him thorough out his matured philosophy of mathematics. We find several examples of this in his intermediary masterwork, the *Big Typescript*:

> What does one *see* "1, 1+1, 1+1+1, …" as?
> As an inexact form of expression. The dots are like additional numerals, but ones that are indistinct. As if one stopped writing down numerals because to be sure one can't write them all down. , but as if they were all there in a kind of box.
> The dots in "1 + 1 + 1 + 1 …." are nothing but four little dots: a sign for which it must be possible to state certain rules. (Namely, the same rules as for the sign "etc, *ad inf*.") This sign does imitate an enumeration in a way, but is isn't an enumeration. And that most likely the rules that apply to it agree up to a point, but not completely, with thoes that apply to an enumeration. (WITTGENSTEIN, 2005, pp. 260, 257-8)

Even as late as in the *Philosophical Investigations* we find him saying:

10 The infinity ingredient becomes apparent if we recast (2) as

$$1 \div 7 = \sum_{n=1}^{n=\infty} \frac{142857}{(1000000)^n}.$$

We will talk about the generality component later.

11 Cf. (WITTGENSTEIN, 1975, pp. 223-34), (WITTGENSTEIN, 2005, pp. 466-7), (WITTGENSTEIN, 1979, pp. 183-6) and (WITTGENSTEIN, 1976, pp. 122-30)

> We should distinguish between the "and so on" which is, and the "and so on" which is not, an abbreviated notation. "And so on ad inf." is not such an abbreviation. The fact that we cannot write down all the digits of p is not a human shortcoming, as mathematicians sometimes think. (WITTGENSTEIN, 2005a, pp. § 208, p 71)

Wittgenstein's worries about the elision dot's notation are clearly *semantical*. In his words, this notation "imitates an enumeration", but it is clearly not one. It's rules agree to a list only "up to a point". So, what are exactly the rules that govern it? To understand better Wittgenstein's point here it is best to go back to Frege's ideas on identity statements (and also to Kripke's ideas about "non-standard" continuation of infinite series). To anticipate a little, Wittgenstein complains that the identity such as (2) is *ill formed!* But let us follow his arguments step by step.

"$1 \div 7 = 0.142\ldots$" is an identity. So, according to (FREGE, 1977, p. 43) one should find nominal expressions on both sides of the relational predicate "$=$". "$1 \div 7$" is surely a fregean name (albeit a composed one), but what about "$0.142\ldots$"? Is "$0.142\ldots$" a *singular term*? Once again invoking Frege, we could say: if "$0.142\ldots$" is to function as a singular term, it has to succeed in denoting *one* and *only one* object. According to Frege, existence, the first requirement for a nominal expression is not so essential. In a famous passage of *On Sense and Reference*, he is notoriously willing to accept "*Odysseus*" as a *name*, even though it has no reference (FREGE, 1977, p. 62). But as to the second requirement, *unicity* of denotation, he was never willing to relax it. If a referring expression is to function as a *singular term*, it has to denote *at most one single object*.

It is Frege's second, more crucial requirement, that, according to Wittgenstein, "$0.142\ldots$" is not capable of satisfying. To see this one has only to consider some "non-standard" possible continuations for "$0.142\ldots$" (KRIPKE, 1972, p. 16). The standard continuation would, of course, be "$0.142\mathbf{857142}\ldots$". But, what about alternative continuations such as "$0.142\mathbf{1}142\mathbf{1}24\ldots$" or "$0.142\mathbf{2}142\mathbf{2}24\ldots$"? Which of them is the (correct?) continuation of "$0{,}142\ldots$"? This is important when we consider the following identities:

$$1241 \div 9999 = 0.124\ldots$$
$$1242 \div 9999 = 0.124\ldots$$
$$1243 \div 9999 = 0.124\ldots$$

Would all these statements be *true*? All of them *simultaneously true* (together with (2))? And so, by transitivity, should we conclude that:

$$1241 \div 9999 = 1242 \div 9999 = 1243 \div 9999 = 1 \div 7?$$

Wittgenstein's worries are clear: "0,142 ..." could *not possibly be a singular term*, a numeral, just because if there were such a term, *it would (ambiguously) denote a infinitude of different numbers!* In fact the most reasonable course would be simply to abandon the idea of "0.142 ..." as a *singular term*. According to this later option, "0.142 ..." *would not be a numeral*, "0.142 ..." would be *a general term!* It's meaning could be elucidated as something like:

"*any infinite decimal fraction beginning by the digits '0', '1', '2' and '4'*"

Of course if we were to reclassify 0.142 ..." as a general term, another difficulty would soon come up. What would a *general term* be doing flanking an identity statement such as (2)? Once again we are back against fregean requirements. What would be the truth conditions of these ("strange") identities? As we anticipated before Wittgenstein concludes that we should simply discard all such statements as being *ill-formed*.

The elision dot notation has elude us with its idea of some "*additional, hidden digits*" hiding in the infinite continuation of "0.142 ..." which would fix the correct intended denotation, finally transforming it into a singular term, a numeral. The philosopher writes:

> The incorrect conception of the word "infinite", and of the role of "infinite expansions" in the arithmetic of the real numbers, seduces us into thinking that there is a uniform notation for irrational numbers (namely the notation of the infinite extension, e.g. of infinite decimal fractions) (WITTGENSTEIN, 2005, p. 498)

Thus, the passage we quoted above, from the *Philosophical Investigations*: "*We should distinguish between the "and so on" which is, and the "and so on" which is not, an abbreviated notation.*" "0.142 ..." is not an abbreviated notation. It is not even a numeral. And, as we will see, we should reject "0.142 ..." as a numeral, not because of any *epistemological* qualms about infinity, such as the Constructivist's, but merely due to semantical requirements, the necessity of clearly maintaining the distinction between *general and singular terms!*

The missing operation

Let us return to our discussion of identities involving recurring decimals, of an identity such as (2) $1 \div 7 = 0.142\ldots$. There is an obvious remedy for our difficulty: we are here dealing with a recurring decimal. There is a cycle involved. So in this case we could use this cycle to finally fix our notation and obtain a singular term.[12] There are many ways to do that. A particularly simple one is the traditional "cycle dot notation", used in some analysis' books:

(2') $$1 \div 7 = 0.\dot{1}42857\dot{}$$

For Wittgenstein, differently from statements such as (2), (2') is obviously well formed. But just as in these analysis' books that employ the cycle dot notation, it is very important for him, as it is for them (GARDINER, 1982, p. 67), to point out that numerals such as $0.\dot{1}42857\dot{}$ have very different mathematical properties than those of ordinary numerals such as "0.142857". A well known example would of course be:

$$0.\dot{9} = 1$$

12 Any mathematician is bound to be worried. Our "remedy" has clear revisionary implications here. How would we deal with *all* the other (much more "important") cases? Wittgenstein's proposals *are* clearly revisionary, there is no doubt about that. They *do not* "leave mathematics as it is". But let us stay calm and follow his proposals step by step. There are many further elements that have to be drawn into our scenario.

Wittgenstein stresses the differences in the "grammars" of these different kinds of numbers:

> The confusion in the conception of the "actual infinite" arises from the unclear concept of "irrational number", that is, form the fact that constructs that are logically quite different are called "irrational numbers" without any clear limits being given to the concept. (WITTGENSTEIN, 2005, p. 496)
>
> The discovery of the periodicity is really the construction of a new symbol and a new calculus.
>
> $1/3$ = "$0.\dot{3}$" is not the same kind of thing as "$1/2 = 0.5$";
>
> "$0.\dot{3}$" is not the result of a division (quotient) in the same sense as 0.375. (WITTGENSTEIN, 2005, p. 446)
>
> One can say of the sign "$0.\dot{3}$ that it is not an abbreviation. (WITTGENSTEIN, 2005, pp. 446, 451)

So, even though for him, say, "$1 \div 3 = 0.\dot{3}$" is perfectly all right, it is a very different type of statement from "$1 \div 2 = 0.5$". But what would exactly *be* this difference? How *in his opinion* should we construe the demarcation between these two identities? Wittgenstein's answer to this question is once again surprising. As an initial bold approximation, we could say that he basically claims that "$1 \div 2 = 0.5$" is a *singular statement* whether "$1 \div 3 = 0.\dot{3}$" is a *general* one! At (2005, pg. 472), for example, we find him comparing many alternative notations for a recurring identity such as "$1 \div 3 = 0.\dot{3}$" to general statements such as addition's *associative law!* A little before that he writes:

> On the other hand the generality of this rule [$a + (b + c) = (a + b) + c$] is none other than that of the periodic division $\frac{1}{3} = 0.3$. That means that nothing in the rule is left open or is in need of completion or the like. (WITTGENSTEIN, 2005, p. 469)

Even thought extravagant, Wittgenstein's line of reasoning here is not very difficult to follow. One of the alternative notations he introduces, while discussing recurrence is:

$$\overset{1}{\mathbf{1}/_{\mathbf{3}}} = \mathbf{0,3}, \qquad \overset{2}{\mathbf{1}/_{\mathbf{3}}} = \mathbf{0,33}, \qquad \overset{3}{\mathbf{1}/_{\mathbf{3}}} = \mathbf{0,333}$$

(2005, pg. 471)

But these are "partial" results. How could we express, employing Wittgenstein's new notation, what was meant in a "non-partial" statement such as "$1 \div 3 = 0.\dot{3}$" or "$1 \div 7 = 0.\dot{1}42857\dot{}$"? In the latter case, the left side of our equation would of course simply be

$$\mathbf{1}/_{\mathbf{7}}^{\mathbf{n}}$$

But what about the new right side of (2)? What should we write down?

$$\mathbf{1}/_{\mathbf{7}}^{\mathbf{n}} = ???$$

According to Wittgenstein[13], our difficulty is a consequence of an *oversight*. We have missed an *operation hidden on the right side* of a statement such as (2). The "*hidden operation*" the philosopher has in mind here would be of course the operation of *repetition*. In the case of (2) we would have something like:

$$\left[\lambda n.\, \mathrm{Rep}["142857"]_{(n)}\right]$$

the operation of "*repeating the string of digits '142857'* ***n*** *times*".

If we use now the symbol "$\frown$" for "concatenation", we could finally obtain the desired *general term* which should occupy the right side of (2):

$$0\frown.\frown \mathrm{Rep}["142857"]_{(n)}$$

We could finally complete our rewriting of (2) as:

(2'') $$\mathrm{Div}[1,7]_{(n)} = 0\frown.\frown \mathrm{Rep}["142857"]_{(n-1)}$$

13 Or our reconstruction of it

Just as Wittgenstein said, in (2'') we now surely have a *very different statement* from a *singular* enunciation such as "'$1 \div 2 = 0.5$". And just as he said, we can now clearly see the use of variables, just like as in the statement "$a + (b + c) = (a + b) + c$". They are both *general statements*.

From a traditional perspective, it would seem that we would only have to make one final logical gadgetry explicit, the *universal quantifier*. The variable "n" is of course free in (2'). if we link it, we would get:

(3) $\forall n \left(\mathrm{Div}[1,7]_{(n)} = 0 \frown . \frown \mathrm{Rep}["142857"]_{(\mathrm{n}-1)}\right)$

Would that be it? Would (3) be a completely acceptable version of the ordinary "$\mathbf{1 \div 7 = 0.142}$ ...", even for finicky Wittgenstein? Was that all he had to say about (the rather plain) recurring decimals?

As anyone that is familiar with Wittgenstein's philosophy of mathematics would have guessed, the philosopher would still not be satisfied. Should we take the universal quantifier in (3) *classically* or, say, *intuitionistically*? Should we understand (3) as:

(3') $\forall_{\boldsymbol{classic}} n \left(\mathrm{Div}[1,7]_{(n)} = 0 \frown . \frown \mathrm{Rep}["142857"]_{(\mathrm{n}/6-1)}\right)$

or as

(3'') $\forall_{\boldsymbol{intuitionistic}} n \left(\mathrm{Div}[1,7]_{(n)} = 0 \frown . \frown \mathrm{Rep}["142857"]_{(\mathrm{n}/6-1)}\right)$?

These differences do matter, of course. If we accept the *classic calculus* and it's *law of interdefinability of the existential*, (3') would turn out to be equivalent to:

(3') $\neg\exists_{\boldsymbol{classic}} n \left(\mathrm{Div}[1,7]_{(n)} \neq 0 \frown . \frown \mathrm{Rep}["142857"]_{(\mathrm{n}/6-1)}\right)$

Not so, of course, if we understand intuitionistically the quantifier. How should we understand the generality involved here? Are these the only two options available? Would there be a third alternative to construe that generality? Following Wittgenstein lead, Dummett

has always emphasized the idea that those logical options run very deep: they involve very basic questions regarding the very semantics of mathematical statements.

> The answer to the question how it is possible to call a basic logical law in doubt is that, underlying the disagreement about logic there is a yet more fundamental disagreement about the correct model of meaning, that is, about what we should regard as constituting an understanding of the statement. (DUMMETT, 1991, p. 17)

Decisions on the rules of logic concerning mathematical statements should be derived from much more general decisions about their semantical interpretation. In other words, we are back to the problem of how to interpret identities and, more generally, how to interpret mathematical statements. Questions about what do they really *assert*, what should we take as being their *content*, their connection to "reality", etc. We are thrown right into deep philosophical water; there is no escaping from that.

III – Mathematical Facts

As we've said before, this paper's main objective is to present Wittgenstein treatment of mathematical identities as back and forth correction. But before we finally do that, though, we will have to quickly review some central elements of the classical and the intuitionistic philosophical construal of mathematical statements in general. These will be important because, by way of contrast, we hope it will end up shedding light on Wittgenstein's proposals.

We will start from a very general interpretational principle about mathematical statements over which *the classicists and the intuitionists are in agreement*. For both these approaches mathematical statements are *true*. And they *both* understand this "being true" as some sort of *correspondence*: a statement is true because there is *something in* reality that *makes it true*. As we will see, classicists *and even intuitionists* accept the idea of "*mathematical facts*" habiting some "*abstract mathematical reality*". It is the (existence) of *these facts* that

turn our mathematical propositions into *true* mathematical statements.

A classical handling of recurring decimal's identities

The idea that *truth* of mathematical statements is somehow a consequence of the *existence of* "*mathematical facts*" (in some "*platonic realm*") is usually taken as *the distinctive trait* of the classical, platonistic approach. This is not a good characterization of the classical position though. As we already anticipated, it would assimilate contemporary intuitionism into the classical camp. This is particularly clear if we take into account the recent proposals, made by Dummett and Prawitz, for distinguishing "*warrant assertability*" from "*truth*". But let's not anticipate too much and begin step by step with the classical approach.

Let us consider once more the algebraic version for a recurring decimal's equation we attributed to Wittgenstein, with its two operations, *division* and *repetition*:

(2'') $\mathrm{Div}[1,7]_{(n)} \;=\; 0 \frown . \frown \mathrm{Rep}["142857"]_{(n-1)}$

In it we have two "unary operations": the operation of "*dividing* $1 \div 7$ *to the* n^{th} *decimal place*", $\left[\lambda n.\, \mathrm{Div}[1,7]_{(n)}\right]$, and the operation of "*repeating the digits* $'142857'$ m *times*", $\left[\lambda n.\, \mathrm{Rep}["142857"]_{(m)}\right]$. The key mathematical content of (2'') would depend on the fact that, for all decimal places, these two operations coincide.

For a classicist, both these operations should be understood as *functions*, i.e., as (infinite) *sets of ordered pairs* (of natural numbers). So, according to classical set theory, even if we accept Wittgenstein's proposal (2''), we would still understand it classically as two *different names* denoting one and the same *mathematical object*. According to this approach, this "key mathematical fact" behind (2'') would be the *fact* that both *singular expressions*

$$\text{“}\left\{ (\boldsymbol{n},\boldsymbol{m}), \in \mathbb{N} \;:\; \boldsymbol{m} = \mathbf{Div}[\mathbf{1},\mathbf{7}]_{(\boldsymbol{n})} \right\}\text{”}$$

and

$$\text{“}\left\{ (\boldsymbol{n},\boldsymbol{n}), \boldsymbol{n} \in \mathbb{N} \ : \ \boldsymbol{m} = \mathbf{Rep}\ [\text{"142857"}]_{(\mathbf{n})} \right\}\text{”}$$

denote *one and the same set*, the set:

$$\left\{ (1,\mathbf{1}), (2,\mathbf{4}), (3,\mathbf{2}), (4,\mathbf{8}), (5,\mathbf{5}), (6,\mathbf{7}), \ldots \right\}$$

Wittgenstein's reading of a recurring decimal's identity, with its two operations, is not the usual one, of course. One doesn't normally understand identities involving recurring decimals as *general statements (*with variables), but as *singular statements* involving an infinite object:

(4) $$1 \div 7 = 0.142\ldots$$

Thus, in a more ordinary reading of these identities, *recurrence* would not even be mentioned, we would only have an operation on the right hand side of the equation and an infinitary object on the left, the "result".

Thus, differently from Wittgenstein, recurrence would not be taken as part of the very semantical characterization of "$0{,}142\ldots$", but as a *property of some divisions*. If we understand the operation $\left[\lambda n.\, \mathrm{Div}[1,7]_{(n)}\right]$ classically, as set of ordered pairs:

$$\mathbf{1} \div \mathbf{7} \ = \ \left\{ \ (\boldsymbol{n},\boldsymbol{m}), \in \mathbb{N} \ : \ \boldsymbol{m} = \mathbf{Div}[\mathbf{1},\mathbf{7}]_{(\boldsymbol{n})} \ \right\}$$

we could then take "*repetition*" as the property of that set. More precisely, if we understand *recurrence* as

$$[\boldsymbol{\lambda f}.\, \exists \boldsymbol{c} \forall \boldsymbol{n}\ (\ \boldsymbol{d} = \ \boldsymbol{f}(\boldsymbol{n}) \rightarrow \boldsymbol{d} = \boldsymbol{f}(\boldsymbol{n}+\boldsymbol{c})\ )]$$

we would take that assertion as:

(4') $$\left|\ \boldsymbol{\lambda f}.\, \forall \boldsymbol{n}\ (\ \boldsymbol{d} = \ \boldsymbol{f}(\boldsymbol{n}) \rightarrow \boldsymbol{d} = \boldsymbol{f}(\boldsymbol{n}+\mathbf{6})\ )\ \right|\left(\ \mathbf{1} \div \mathbf{7}\ \right)$$

Her we would be talking about an (*abstract*) *object*, the "*operation* $1 \div 7$", and would be asserting of that object that it *really obeys a 6-cycle*. This would be the "mathematical fact" involved . It is because

the infinite object "**1 ÷ 7**" *really has the property of recurrence* that (4') happens to be a *true* mathematical statement.

Intuitionistic mathematical facts

Contrary to a rather common misconception, the adherence to the *correspondence theory* of mathematical truth, to the idea of *mathematical facts*, is not characteristic only of the classical approach. From the very beginning in 1959, in his famous article *Truth*, Dummett has emphasized a similar intuition:

> …the correspondence theory expresses one important feature of the concept of truth which is not expressed by the law "it is true that *p* if and only if *p*" …: that a statement is true only if there is something in the world *in virtue of which it is true*. (DUMMETT, 1978, p. 14) *apud* (MARTIN-LÖF, 1995)

Mathematical statements are *descriptive*. They record (when true) mathematical *facts*, aspects of a *purely mathematical reality*. And of course it is this fundamental representational property that distinguishes *true* mathematical statements from *false* ones. Dummett calls this philosophical proposal, the idea of *correspondence*, principle "*Principle C*".

Several years later, in another famous article from 1976, *What is a Theory of Meaning II?*, Dummett better qualifies his adherence to correspondence. He concedes that, taken in isolation, Principle C is rather "empty" and adds that its main role is to "settle on the appropriate notion of *truth* for various types of statements":

> In general, we can learn something by applying the principle C to a specific type of statement only when we have already decided something about the sort of thing in virtue of which a statement of that type can be true; (DUMMETT, 1993, p. 53)

So, in the case of mathematics, it becomes natural to ask what type of "*facts*" should we take to be behind the truth of mathematical statements. What kind of *entity* should we elect as the *truth-makers* of mathematical propositions? For a long while, the official intui-

tionistic answer to this question was the very well known and direct:

> Intuitionistically, truth of a proposition is analyzed as existence of a proof; a proposition is true if there exists a proof of it. (MARTIN-LÖF, 1991, p. 141)

Very recently though, in the late nineteennities, there was a dramatic shift in the way the leading intuitionists answered this question with two *different* proposals, made by Dummett (DUMMETT, 1998) and by Prawitz (PRAWITZ, 1998), for a distinction between the notions of "*warrented assertion*" (i.e., possession of a proof) and "*mathematical truth*". For both Prawitz and Dummett, the "*mathematical fact*" that turns intuitionistic statements into true mathematical propositions should not be taken to be the simple *existence of a proof* anymore, but "*something else*". They agreed that the existence of the proof would only *assure* us that this fact exists, but should not be understood as coinciding with it (PRAWITZ, 1998, p. 46). Still, they *diverged* on what this "something else" should be.

The recent shift on what the intuitionists regard as the truth maker of propositions should not come as a surprise. It was already concealed in the simplicity of the previous formula: "*truth is the existence of a proof*". All we would have to do is to ask the intuitionist about the concept of *existence* being employed in that formula. Are we talking about "*concrete existence*", "concrete possession of a proof", or merely its "*existence in principle*"? As it is well known, if we pressed on the point, the initial simplicity of the formula would give away to much finer distinctions. The intuitionists work with *two different notions* of *proof*, *canonical* and *non-canonical* proofs[14]. And it turns out that the role played by notion of "*existence*" regarding these two kinds of proof is also different.

The notion of existence pertaining to a *canonical proof* is the hardcore notion of "*concrete empirical existence*". But in the case of a *non-canonical proof*, the situation is different. A non-canonical proof

14 Dummett also calls them "proofs" and "demonstrations" (DUMMETT, 1991, p. 177).

expresses only *existence "in principle"*. As we already anticipated above, a proof like that enunciates "*the possibility of obtaining* a canonical proof" from it. So a non canonical proof is really only a method that, if it *were applied, could generate* the desired *canonical proof* (PRAWITZ, 1977, pp. 22, 26; MARTIN-LÖF, 1987, p. 413). In other words, it asserts that *that possibility really exists,* guaranteed by the (concrete) existence of an *intensional object*, the method.

The central notion of "*proof*" is thus clearly that of a *canonical proof*.[15] It is the existence of those proofs as *syntactical events* that could be identified with "mathematical truths". But Intuitionism does not restrict itself to that notion of proof

> ...It is not true, even Intuitionistically that the condition of asserting a sentence is that we know a proof of it in this sense [the canonical sense]. (PRAWITZ, 1977, p. 21)

To be able to rise above the recording of trivial mathematical identities, Intuitionism is force to include also a second notion of proof, that of a non-canonical demonstration. So, our question now is: what would be the truth-maker for this second type of proof, *non-canonical proofs*?

There are two possible answers here. We could chose the *method,* which indirectly ensures the possibility in principle of obtaining the desired canonical object. Or more directly, we could elect the (existence of) the very *in principle possibility*. Dummett recommends the *method* as truth-maker of mathematical propositions. (DUMMETT, 1998, p. 123)[16] Prawitz, on the other hand, insists that that notion of "method" would still be *temporal*: it is restricted to the methods *contemporarily at our disposal*. So he proposes to go further and accept a stronger, *atemporal notion,* of "*potential existence* of a proof" (and of a method). (PRAWITZ, 1998, p. 48). Future

15 In the case of natural numbers, for example, the canonical notation for them has some very important properties: "$ssssss0$" both *denotes* the number 6 but also concretely *instantiates* six successors, "$ssssss$".

16 It is important to notice that this notion of *method* should not be identified with the classical notion of *computability* on pains of an infinite regress (PRAWITZ, 1998). It has to be taken as primitive.

(presently unknown) methods and canonical proofs would have a "potential existence", even today! Either way we have abstract realm populated by "intensional objects", "methods", in the case of Dummett, or populated by "potentialities", in the case of Prawitz.

We won't go into to the problems these proposals may have. Even the more sober one, advocated by Dummett, seems strange at times. For example, we do have a method for calculating Pi's decimal places. But we haven't calculated them all. What should an intuitionist say about these *uncalculated decimal places* (even if he sides with Dummett)? We would have to say that they *exist* "in some sense", for their existence is guaranteed by that method. But we don't really *have them*, because we haven't calculated that far (and never will, for some of these decimal places). So, what should we answer here? What about the decimal places not only as yet uncalculated, but so enormous that it would take, say, the number of seconds since the big band to calculate? Should we still insist that they exist because they "*could be reached*"? What kind of "capability" would that be? Here we seem to be flirting again with the dangerous concept of an "*impossible possibility*".

Let us see now Dummett's handling of this problem. We do have a method for calculating Pi, so his answer could not be really much different from:

> It seems that we ought to interpose between the platonist and the constructivist picture [a reference to Wittgenstein] an intermediate picture, say, of objects springing into being in response to our probing. We do not *make* the objects but must accept them as we find them (this corresponds to the proof imposing itself on us); but they were not already there for our statements to be true of false of before we carried out the investigations which brought them into being. (DUMMETT, 1978, p. 185)

But shouldn't we still ask: what crucial difference could there be between saying (with the platonist) that these objects *do* exist, and Dummett's insistence on the notion of objects "*springing into being in response to probing*"? Wouldn't that be just a play of words or should we take his proposal more seriously? But if we do, should we also

accept, say, some kind of "*berkelean cake*" that would "*spring into being*" every time we open our refrigerator door?

"Two facts" idea

In the last section we've seen that *both* the classical and the intuitionist accept a notion of "*mathematical facts*". By accepting this notion, it becomes natural for both cases do understand the idea of "*applied mathematics*" as *instantiation*. An application of mathematics would be an instantiation of an *abstract mathematical structure (form)* onto, say, a *physical* one. The details of how this instantiation would take place are different though for classicists and intuitionists. The very truth makers involved are distinct.

Let us take first the better known classical approach. In Analysis, for example, we have the purely abstract concept of *real numbers*. And there are also many *mathematical facts* about them. We can then geometrically *apply* this theory by using the famous picture of the real numbers as *points on a line*. Through this picture, *numbers* become *names* of *positions* along a geometrical line. And we can further take this line as, say, a concrete dimension of a *physical extension*. Our numbers become then *names* of *concrete positions in space*. And, of course, *mathematical facts* about these numbers become *physical facts* about this (concrete) space.

How would an intuitionist view the idea of *applied* mathematics? The key point here is the idea that intuitionistic mathematics is made, not of *propositions,* but of *judgments*. And judgments are *acts*. (MARTIN-LÖF, 1987, p. 417) This is why it is so natural for an intuitionist to view *programs executed by computers* as the very embodiment of mathematics (MARTIN-LÖF, 1979, pp. 5-6). Still, even in the case of computers, we should differentiate the *abstract program* a computer is supposed to be executing from the *concrete machine* in front of us. But then, when would we say that the machine is *really executing* a particular (abstract) program? Well, a program is a *specification* of how that equipment is *supposed to behave*, if it is to instantiate that algorithm. In other words, we say that the machine is *really following that program* when the space of possible *concrete*

behaviors of the machine exactly *reproduces* the *abstract* space of behaviors prescribed by it.

Once again, we have the idea of an application as an *instantiation*. The *abstract* program is *instantiated* onto the *concrete* machine. And, just as before, we can distinguish *two kinds of facts* in this application. We have the "*mathematical fact*", that that method produces that result. And we have the "*physical fact*" that that concrete equipment in front of us is (really) following that specification, that the *mathematical facts* prescribed by the program are really being instantiated by the behavior of that machine.[17]

Wittgenstein was always critical of the "*two facts idea*". In a lecture in 1939 he directly says:

> It is not the case that there are two facts – the physical fact that if one counts the squares [of a rectangle 36 squares long and 21 squares wide] one gets 756 and the mathematical fact that 21 times 36 equals 756. (WITTGENSTEIN, 1976, p. 39)

Not only Wittgenstein refuses the idea of abstract mathematical *objects,* such as the *numbers* and *sets* as proposed by the *classic* but, more importantly, he refuses also the *intuitionist* postulation of "*abstract possibilities*" as opposed to "*real, physical possibilities*".

> We imagine possible structures and impossible ones, and we distinguish both from real structures. Is seems as though in mathematics we showed what structures are conceivable, imaginable, not real. (WITTGENSTEIN, 1976, pp. 145-6)

He was always critical of the idea of *abstract mathematical possibilities* as "*shadows*" of *concrete, real possibilities*.

> We say, for example, that a machine *has* (*possesses*) such-and-such possibilities of movement; we speak of the ideally rigid machine which *can*

17 In the end, Prawitz seems to have the right intuition. The *truth maker* for intuitionist mathematics appears to be the notion of *possibility*. It would be odd to say that the possibility exists *because* the method exists. It seems more natural to say that the method *really* works because it describes something *really possible* (an "effective" way of generating that result).

> only move in such-and-such a way.– What is this *possibility* of movement? It is not the *movement*, but it does not seem to be the mere physical *conditions* for moving either …The possibility of a movement is, rather, a shadow of reality. (WITTGENSTEIN, 1983, p. 86)
> *The difficulty here is to defend oneself against the thought that possibility is a kind of shadow of reality.*
> *It is one of the most deep-rooted mistakes of philosophy to see possibility as a shadow of reality.* (WITTGENSTEIN, 2005, pp. 258,259)

Real, *Logical* and *in principle* possibilities

We have here a key point that neatly differentiates Wittgenstein's and both the *classical* and the *intuitionist's* approaches. For both these latter approaches, besides the modal notion of *real possibility (*understood counterfactually as situations that *might happen to occur*), we should accept *other modal notions*. As we've seen, in the case of the intuitionist, we have the notion of "*in principle* possibility". An "in principle possibility" would be a kind of "*second level* possibility": a *possible existence* of a *real possibility*.

The classicist postulates a third notion of possibility, as distinct from both the intuitionist, *in principle,* one and from *real possibility*. It is the notion of a "*classical logical-mathematical possibility*". The contrast between this notion and the intuitionist's *in* principle one becomes apparent when we consider mathematical possibilities that not only *happen* not to describe real possibilities but *never could*, *even "theoretically"*. As Charles Parsons has pointed out years ago, this is precisely a novelty within Cantorian Set Theory:

> It is only when higher infinities of Cantorian set theory are introduced that mathematical objects must violate the conditions of representability in concrete terms. … If the "physically possible" is what can in some sense be realized in space and time, then structures of sufficiently high cardinality … are not physically possible. (PARSONS, 1983, p. 191)

No wonder the intuitionists were so critical of Cantor's mathematical hierarchy of infinitudes. Beyond a certain level[18], not only are those mathematical structures *happen contingently to be non-instantiable* (even considering "*theoretical*" possibilities), but are actually *necessarily non-instantiable*.

The neat differentiation between Wittgenstein and both the intuitionists and the classicists becomes clear when we consider the fact that the philosopher refuses, not only the stronger classical modality, but even the *second level*, *in principle,* notion advocated by the intuitionists.

> There is a feeling: "There can't be actuality and possibility in mathematics. Everything is on *one* level. And in fact, is in a certain sense *actual*". (WITTGENSTEIN, 2005, p. 495)

As Raymond Bradley has suggested in his series of equations between key *semantical*, *modal* and *subjective* notions, all the way from the Tractatus Wittgenstein's philosophy can be generally characterized *by the refusal of any notion* of "*abstract possibilities*" as distinct from *"real" ones*. For him, the three notions

(Real) possibility = *In principle possibility* = *Logical possibility*

should always *coincide!* (BRADLEY, 1992, p. 34)

Non-descriptive Mathematics and Predication

Besides these *modal* intuitions, Wittgenstein has two parallel theses regarding the *semantics* of mathematical statements. The first one is that, contrary to the idea of *correspondence* suggested by both the classicists and the intuitionists, for him mathematics *does not describe any reality*.[19]

18 Parsons suggests $2^{2^{\aleph_0}}$. (PARSONS, 1983, p. 191)

19 It can be traced of course all the way to the famous *Grundgedanke* of the Tractatus. (WITTGENSTEIN, 1961, p. 4.032)

> In mathematics *everything* is algorithm, *nothing* meaning; even when it seems there's meaning, because we appear to be speaking *about* mathematical things in *words*. What we're really doing in that case is simply constructing an algorithm with those words. (Big Typescript. pg. 494)
> I am trying to say something like this: even if the proved mathematical proposition seems to point to a reality outside itself, still it is only the expression of acceptance of a new measure (of reality). (RFM, 162)

This is a very early insight of Wittgenstein's. It is already clearly present in the famous *grund gedanke* of the *Tractatus*. (WITTGENSTEIN, 1961, p. 4.032). *Empirical*, *contingent* propositions *describe*. *Necessary* statements such as logical laws, mathematics or even philosophy do not *describe* anything, but *prescribe*.

For Wittgenstein, the confusion between these two "grammars" is a very common, deeply rooted and extremely pernicious mistake among philosophers and mathematicians. He is always adverting against the idea of approximating *empirical* and *mathematical* statements:

> The confusions in these matters are entirely attributable to treating mathematics as a kind of natural science.
> Nothing is more disastrous to philosophical understanding than the notion of proof and experience as two different – yet still comparable – methods of verification. (WITTGENSTEIN, 2005, pp. 429,419)
> Mathematicians, when they begin to philosophize, always make the mistake of overlooking the difference in function between mathematical propositions and non-mathematical propositions.
> These discussions have had one point: to show the essential difference between the uses of mathematical propositions and the uses of non-mathematical propositions which seem exactly analogous to them. (WITTGENSTEIN, 1976, p. 111)

He goes as far as denying that the notions of *truth* and *falsity* should be applied to mathematical statements:

> The terms "sense" and "nonsense", rather than the terms "true" and "false", bring out the relation of mathematical propositions to non-mathematical propositions. (WITTGENSTEIN, 1979, p. 152)

Not surprisingly, the big differences in *semantical function* Wittgenstein sees between *empirical* and *mathematical* statements determine for him parallel fundamental distinctions in the very *grammatical structure* of these assertions. In fact, the philosopher denies that *predication* can be at all present in mathematical statements. One again, the textual evidence is very forceful:

> Is 2 + 2 = 4 a proposition *about* 2 and *about* 4? Compare this proposition with "There are no other men in this room than Jack an John". …if "There are two men here" is taken to be about 2, then it is misleading to say 2 + 2 = 4 is about 2; for it is "about" it in a different sense. (Lectures 1934-35, pg. 155)
> …a mathematical proposition is not *about* its constituents in the sense in which "The sofa is in this room" is *about* the sofa. . (LFM, Lectures 1939 XXVI, pg. 254)
> …mathematical propositions do not treat of numbers. Whereas a proposition like "There are three windows in this room" *does* treat of the number 3. (LFM, Lectures 1939 XXVI, pg. 250)

> Suppose I say "Prince has blue trousers"; that is a proposition about trousers. (….)
> What about "two"? "2 + 2 = 4" – but this *isn't* about 2; it is grammatical.
> *Turing:* Isn't it merely a question of how one extends the use of the word "about"?
> *Wittgenstein*: That is a most important mistake. – Of course you can say mathematical propositions are about numbers. But if you do, you are almost sure to be in a muddle. (LFM, Lectures 1939 XXVI, pg. 251)

IV – Back and Forth Correction
The Metalinguistic Component

We have dedicated the entire part III of our article to review some very general traits on how classicists and intuitionists construe the semantics of mathematical statements and their application, say, to physics. We've seen that both these schools propose nothing less than a bifurcation of the notion of *reality* in two. Aside from the *concrete*, *empirical reality,* we would also have an *abstract, logical-*

mathematical one. Once this bifurcation is effected, we can then explain *mathematical truth* as the existence of the appropriate *facts* within the abstract realm. On the other hand, *application* of mathematics is construed as *instantiation*. There is a isomorphism between some aspect of empirical reality and some mathematical structure. Through this isomorphism, names of mathematical objects (numerals, for example) become also names of empirical entities (say, extensions). And thus mathematical facts concerning these objects are nicely transferred to that part of empirical reality.

As we've seen, Wittgenstein was very critical of the "*two facts*" idea. He rejected altogether any notion of a *mathematical abstract reality*, even if this reality was to be populated only by *abstract potentialities*. For him, mathematical statements neither *talk about*, nor *answer to* any *abstract realm*, no matter how this realm might happen to be construed. In fact, he goes as far as denying the *very applicability* of *predication* (and of the notion of "*truth*") to mathematical statements! Mathematics simply *does not talk about* anything.

If mathematical statements do not *talk about anything*, how is their *semantics* to be understood? If, say, "$2 + 3 = 5$" is not a *statement about numbers,* nor any other *abstract objects*, not even about *abstract possibilities* (of obtaining some canonical form), *what is it about then?!* What semantical role would Wittgenstein *accepts* for mathematics? What do mathematical statements *assert*? After all, in order to have any *meaning*, it appears that they are bound to assert something...about something else. This is where the notion of "*back and forth*" correction comes in. There are two elements involved in this notion. The first, and most important one, is the *metalinguistic element*. The second one is the *normative* and *deontic component*. Let us begin with the metalinguistic component.

As in the case of so many of Wittgenstein's key ideas, this one is also derived from Frege. In his *Grundlagen der Arithmetik*, we find this very striking passage by the German philosopher:

> It is true that at first sight the proposition "All whales are mammals" seems to be not about concepts but about animals; but if we ask which animal then are we speaking of, we are unable to point any one in particular. Even supposing a whale is before us, our proposition still does not state anything about it. We cannot infer form it that the animal

> before us is a mammal without the additional premise that it is a whale, as to which our proposition says nothing. (FREGE, 1978, p. 60)

The image of an immense whale laying directly before us is extravagant. And when we say "*All whales are animals*" we do seem to be *talking about whales* (saying that they are mammals). This is what traditional logic (and ordinary grammars) has always taught us, ever since Aristotle. And there is that enormous animal in front of us. But Frege refuses the idea that the rule "*All whales are animals*" does in any way talk about those concrete creatures. As we know, his strange proposal was that the statement assertes something about a *complex concept* instead, the concept "$[\lambda x. Whale(x) \rightarrow Mammal(x)]$", saying *about that concept* that it is *universally valid*. This is Frege's famous idea of "*second order predication*": *general statements* do not talk directly about *empirical reality*, but only establish *connections* between *concepts*.

We know that for Frege, even though statements like "*All whales are animals*" did not talk about *empirical reality*, they did talk about an "*alternative abstract reality*". In fact, Frege is famous for having gone as far as fully accepting the idea of a "*third realm*" (FREGE, 1977a, p. 17), an independent abstract reality in which these conceptual connections actually "*were the case*". As we've seen, Wittgenstein could not accept this idea of a bifurcation of reality into realms. But there was something in Frege's proposal that interested him. This was the idea of a "*second order connection*" of the rules *vis-à-vis* empirical reality. Mathematical statements did not talk *directly* about reality, but only establish *metalinguistic rules* (Frege's "connections") for *meaningful employment* of these *words* (Frege's "concepts") in *empirical contexts*:

> One might also put it crudely by saying that mathematical propositions containing a certain symbol are rules for the use of that symbol, and that these symbols can then be used in non-mathematical statements. LFM, lecture III, pg. 33
>
> To sum up, I have tried to show that the connection between a mathematical proposition and its application is roughly that between a rule of expression and the expression itself in use. (WITTGENSTEIN, 1976, pp. 33, 47)

The important point to be stressed here is that on Wittgenstein's hands Frege's second order predication becomes fully *metalinguistic*.

To the despair of his famous interlocutor, he is always introducing the operator "*call*" in all his formulations of mathematical statements

> "This is what we do when we perform the process which we *call* "multiplication". 144 is what we *call* "the right result".
> Supposing we do a multiplication: the use of this is that we aren't willing to recognize a rule of multiplication unless it can be got in a particular way. For instance, we do not accept the rule that 1500 × 169 = 18. We should not *call* that a multiplication. (WITTGENSTEIN, 1976, pp. 97, 106)[20]

Differently from the philosopher, his interlocutor wants to leave mere metalinguistic description behind and go back to the idea of *truth*, good old *answerability to a mathematical reality*.

> "Is that supposed to mean that it is equally correct whichever way a person counts, and that anybody can count as he pleases?" - We should presumably not call it "counting" if everyone said the numbers one after the other *anyhow*.
> "Then according to you everybody could continue the series as he likes; and so infer *anyhow*!" In that case we shan't call it "continuing the series" and also presumably not "inference". (WITTGENSTEIN, 1983, pp. 37, 80)

Wittgenstein's answers – "*..we should presumably not call it "counting…", "…in that case we shan't call it 'continuing the series' and also presumably not 'inference'"* – are precisely meant to reintroduce the dreaded *metalinguistic point of view,* blocking the idea of any "answerability to an abstract reality", the recovery of a "descriptive content".

The Normative Component and Disqualifying Criteria

The second element in Wittgenstein's idea of *back and forth correction* is now immediate. If there is to be no remaining "*descriptive content*"

20 My italics

in Wittgenstein's notion of "rule", then we are left with a purely *normative* connection. As usual, Wittgenstein is ready to follow this idea wherever it might lead. He goes as far as proposing:

> Suppose we look at mathematical propositions as commandments, and even utter them as such? "Let 25^2 be 625."
> Can we imagine all mathematical propositions expressed in the imperative? For example: "Let 10×10 be 100". (WITTGENSTEIN, 1983, pp. 271, 276)

Mathematics doesn't describe anything. It does not talk about ideal, abstract entities. It is purely prescriptive. It prescribes criteria for the correct usage of certain terms such as "*line*", "*angle*", in ordinary *empirical situations*, what Wittgenstein calls "*grammatical rules*":

> Geometry isn't the science (natural science) of geometric planes, lines and points, as opposed, say, to some other science whose subject matter is gross physical lines, strips, surfaces, etc. and that states *their* properties. The connection between geometry and propositions of practical life, which are about strips, color boundaries, edges and corners, etc. doesn't consist in geometry's speaking of things similar to what these propositions speak of, although geometry speaks about *ideal* edges and corners, etc.; rather, it consists in the connection between these propositions and their grammar. Applied geometry is the grammar of statements about spatial objects. (WITTGENSTEIN, 2005, p. 391)

The best way to regard these "rules" is to take them, not as *positive determinations*, but as *negative* ones, as *constraints* on empirical statements. In other words, it is best to regard them as establishing criteria for *disqualification* of empirical claims. Let us take a simple example. Suppose we consider the usual geometric procedure for bisecting an angle with a compass. Does it make sense to say that an "*ideally executed*" instance of that procedure would have the "property" of generating two equal angles (but *rough empirical operations* would just "approximate" that abstract ideal)? No, not for Wittgenstein. According to him, the "mathematical content" of that technique is a *connection* between *certain procedures* and the concept "*equal*

angles". This connection is then used to *judge* (and maybe *discard*) empirical allegations.

> If I regard construction as my criterion, I can by no means check the division of angles by measurement. The case is much rather this: if measuring yields a difference, I shall say, the compass is faulty, that was not a straight line, etc. For construction is now my *standard* according to which I judge the quality of a measurement. (WAISMANN, 1979, p. 205)

Other geometrical examples of rules from his intermediary period are:

> The proposition "corresponding angles are equal" means that if they aren't found to be equal when measured I will declare the measurement incorrect; and the "sum of the angles of a triangle is 180 degrees" means that if it doesn't turn out to be 180 degrees when they are measured I will assume there has been a mistake in the measurement. (WITTGENSTEIN, 2005, p. 391)

Back to recurring decimal's identities

We now have all the elements at our disposal to finally complete our discussion of Wittgenstein's favorite example of a mathematical identity, the ones involving recurring decimals. Let us then go back to where we left our discussion of those identities, to the new version proposed by the philosopher in which the two operations of *division* and *repetition* are clearly brought forward.

$$(2'') \qquad \mathrm{Div}[1,7]_{(n)} \quad = \quad 0 \frown . \frown \mathrm{Rep}["142857"]_{(\mathrm{n}-1)}$$

So, how should we then interpret this general equation according to Wittgenstein? Not so with the help with the usual quantifiers, being them intuitionistic or classic, but as *back and forth correction*. In other words, we are going to employ both the *metalinguistic* component introduced by the expression "*Whatever one calls…*', and the *normative* component introduced by the expression "*one must..*". The complete formulation would be something like this:

(5) *Whatever one calls*
"the result of the *operation dividing to the* n^{th} *decimal place*"
one must also *call*
"the result of the operation of writing "0," and of repeating
"142857" through $n-1$ places,
and *vice versa*

Wittgenstein is very explicit about both these components in his later discussions on recurring decimals. Let us start with the *normative* component. In his *Lectures on the Foundations of Mathematics* in 1939 he writes:

> Here I am adopting a new criterion for seeing whether I divide this properly - and that is what is marked by the word "must". (WITTGENSTEIN, 1976, p. 129)

For Wittgenstein, the mathematical advance involved in the notion of "*recurring decimals*" is best viewed as *mutation of our language*. We find ourselves adopting new criteria for old words.

> We actually have in $\dot{3}$ a new symbol. $\boxed{1_1 : 3}$ *is a new operation and has the result in a different sense than* $1/3$. *(AWL, p 211)*

Before our "*language mutation*" we could actually detect the *repetition* of some digits. But we lacked altogether the concept of "*recursion*".

> If we follow the rules as we do follow them, being prepared as we are, then this is what will always happen [the recursion]. (WITTGENSTEIN, 1976, p. 129)

After the linguistic mutation, we have a new *normative criterion*, expressed by a "*timeless must*", a rule*:*

> Then later he takes recurrence as the criterion: "it must happen".(Timeless "must') (WITTGENSTEIN, 1976, p. 129)

Wittgenstein is not always completely careful in his usage of the two words, "repetition" and "recurrence". In the second quote

above about the linguistic situation before the language mutation, he uses "*recursion*" to refer to the *repetition* of digits one could detect even before the adoption of the new concept of recurring decimals.[21] But even with these minor terminological oscillations, the distinction between *recursion* and *repetition* is clearly emphasized:

> Periodicity does not mean the same as several repetitions of the same number or numbers, but makes a new calculus. (WITTGENSTEIN, 1979, p. 187)
> Similarly with "[I] believe it will recur". (…) You might believe in two totally different things. The phrase is misleading: "will recur" as normally used is a mathematical phrase. It is not a temporal expression; it doesn't mean "will recur [repeat] with most people" or "will recur [repeat] in half an hour" or anything like that. (WITTGENSTEIN, 1976, p. 124)

The connection between "repeating" and "dividing" and new kinds of errors

Let us now focus our attention to the second component in the idea of back and forth correction, the metalinguistic connection between the notion of "*repeating*" and that of "*dividing*". Here the importance of Wittgenstein insistence on a "hidden operation", the "geometrical" operation of "repeating a pattern", becomes clear:

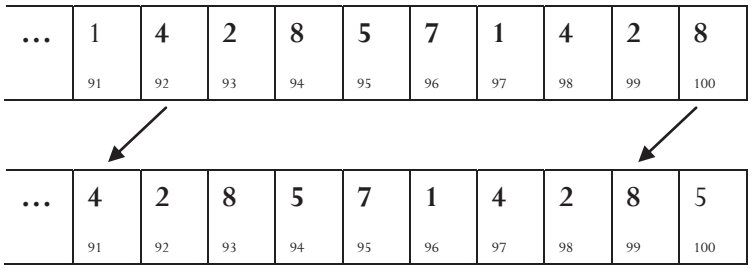
> The question of recurrence is then a strictly geometrical question: the man will be persuaded that if the repeats this pattern here, there must be the same numeral repeated (A new criterion *that he has done so-and-so*) (WITTGENSTEIN, 1976, p. 130)

The new rule connects the two concepts, using one as correcting criteria for the other. Wittgenstein's proposal is best visualized by an imaginary experimental psychology like procedure involving two subjects, A and B. We ask A to perform the *division* $1 \div 7$, let us say, all the way up to 100 decimal places (we can imagine him

21 This may have to of course do with the origin of the text: a transcription of his lectures.

using a Turing machine tape for that). And we ask B to execute the operation of writing "0" and then *repeating* the digits "142857" through 99 extra places (he can also do that in a squared tape). The idea is that we can use the comparison of the two tapes as a way of checking if the two orders were correctly executed. Any mismatch between them would be an indication of a misapplication of (at least) one of the procedures.

It is important to notice that the correcting routine can go either way. Normally, if we've noticed a mismatch between the tapes, it would be reasonable to expect a mistake in A's execution of the command (the one that was carrying out the division). This is due to the fact that his operation is much more error prone than the other. But, if the two tapes happen to be off by, say, only one place and present otherwise identical patterns, such as in:



it would be reasonable to expect a mistake by B (the one carrying out the repetition). This would be because in the case of the division, of course, a mistake in one quotient digit would clearly disrupt the entire recurring pattern (and we could even end up halting at the end of the process).

Wittgenstein is again very conscious of symmetrical character metalinguistic dependence asserted. In his *Remarks on the Foundations of Mathematics* he writes:

> Suppose that when we worked out a division and it did not lead to the same result as the copying of its period. That might arise e.g. from our altering our tables, without our being aware of it. (Though it

> might also arise from our copying in a different way) (WITTGENSTEIN, 1983, p. 263)

It is important to stress this point because, as we anticipated before, it indicates a very significant connection between him and Frege. Even though Wittgenstein is very far from the *triangular* construal of mathematical identities and its idea of an "abstract object" (being denoted by both sides of the equation), his reading of "$1 \div 7 = 0.\dot{1}42857\dot{1}$" is completely *flat*. The correction he proposes *can* go either way (although it might go one way more often than the other). The paring of the two concepts puts them on the *same level*, each one as furnishing correcting criteria of the other.

There is a second, very important point to be emphasized here. Each *new rule* introduces new *kinds of mistakes*. Before the rule connecting the two operations, one could err the execution of the division $1 \div 7$, say, because one of the partial quotients was wrong. After the introduction of the new rule, one could claim that the operation was not actually performed, not because one had any local mistake one could spot, but merely because, say, a digit "6" had showed up within that expansion. It is an essential element in Wittgenstein's view of mathematics that each new *proof*, each new conceptual *connection*, establishes new possibilities of *error*, new ways of *disqualifying allegations* that, say, such as such operations *were performed*.

Each side operation taken in isolation

In the last section we have introduced an imaginary situation in which two subjects, A and B, executed two operations, repeating and dividing, which were connected by a prescription, a rule. It is important to emphasize the purely normative, metalinguistic connection between the rule and those empirical operations. For Wittgenstein, a mathematical rule is completely *impersonal* and *atemporal*.[22] It does not *talk about* any specific empirical operations

22 Once again, Wittgenstein is reminiscent of Frege. Cf. (FREGE, 1964, pp. 12-3)

performed by particular computing agents ("that computer on my desk top") on any specific occasions ("this morning"). In fact, as we've seen before, rules do *not talk about* anything. They merely lay criteria that are used to *disqualify empirical claims*, allegations that, say, such and such an operation was actually performed by some agent in some determinate situation.

Wittgenstein is quite forceful about both the *atemporal* and the *impersonal* character of rules. He writes:

> Questions of fact always involve time; mathematical facts or propositions do not. (WITTGENSTEIN, 1979, p. 184)
> "The 100 apples in this box consist of 50 and 50" - here the non-temporal character of "consist" is important. For it doesn't mean that *now*, or just for a time, they consist of 50 and 50. (WITTGENSTEIN, 1983, p. 74)
> In mathematics we have propositions which contain the same symbols as, for example, "Write down the integral of…", etc., with the difference that when we have a mathematical proposition time doesn't enter into it and in the other it does. (WITTGENSTEIN, 1976, p. 34)

Regarding the *impersonal* trait, contrary to any communitarist reading of rules, Wittgenstein writes:

> "The rule, applied to these numbers, yields those" might mean: the expression of the rule, applied to a human being, makes him produce those numbers from these. One feels, quite rightly, that that would *not* be a mathematical proposition. (WITTGENSTEIN, 1983, p. 228)
> "But mathematical truth is independent of whether human beings know it or not!"--Certainly, the propositions "Human beings believe that twice two is four" and "Twice two is four" do not mean the same. The latter is a mathematical proposition; the other, if it makes sense at all, may perhaps mean: human beings have *arrived* at the mathematical proposition. The two propositions have entirely different *uses*. (WITTGENSTEIN, 2005a, pp. 192-3)

Even if an agent has executed an operation for the first time, for example, a computer spitting out a new place in Pi's decimal expansion, it's results can be taken *atemporally*, that is, if this "new rule" is ever used to "judge proceedings":

> In a calculation I surely wanted from the beginning to know what the result was going to be; *that* was what I was interested in. I am, after all, curious about the result. Not, however, as what I am *going* to say, but as what I *ought* to say. (WITTGENSTEIN, 1983, p. 195)

The importance of stressing the impersonal and atemporal character of rules comes out very clearly if we consider the shift that takes place when we focus on *one of the two operations* paired by the rule *considered in isolation from the other*. Let us take, for example, the operation $[\lambda n.\, Div[1,7\}_n]$. If Wittgenstein accepts a notion of a "*connecting rule*", above and beyond any actual practice of human (and electronic) agents, does he also accepts the notion of an *abstract operation* as a *mathematical entity* distinct from any particular empirical implementation? In other words, does he accepts the notion of "*function*" (and that of an "*algorithm*") as an *abstract specification of a "purely mathematical" procedure*?

Once again, Wittgenstein's proposal is extravagant. His answer is a resounding: No! One can have a rule prescribing converse correctibility of *pairs of* concepts (such as *repeating* and *dividing*). And this pairing is laid down both atemporally and impersonally. But that doesn't mean that we can now, say, "detach" the concept of "*division*" from that rule and talk about an "abstract" version of that operation. We do have the rule

$$\mathrm{Div}[1,7]_{(n)} \quad = \quad 0 \frown . \frown \mathrm{Rep}["142857"]_{(\mathrm{n}-1)}$$

and this rule prescribes atemporally and impersonally. But for him it does not make sense to talk about an "atemporal" and "impersonal" entity represented by each side of that equation taken in isolation. In other words, it does not make sense to talk about some "abstract operation" of division

$$[\lambda n.\, Div(1,7)_n]$$

Wittgenstein writes:

> We use the expression: "the steps are determined by the formula…". How is it used? - We may perhaps refer to the fact that *people* are

> *brought up by their education* (training) so to use the formula $y = x^2$, that they all work out the same value for y when they substitute the same number for x.
> The way the formula is meant determines which steps are to be taken. What is the criterion for the way the formula is meant? Presumably the way *we* always use it, the way *we were taught to use it*. (WITTGENSTEIN, 1983, pp. 35, 36)[23]

It is *here* that the communitarist element in his thinking forcefully comes in. To the despair of the mathematician, Wittgenstein insists introducing "*people*" and the "way *they are trained*" ("nowadays") in his discussion of what would normally be taken as a "purely mathematical" operation.

This is not, of course, what his interlocutor has in mind. The mathematician wants to talk about an *abstract* concept of a "*function*", not a behavior of concrete, calculating agents. Instead of the purely *normative*, *metalinguistic* connection (between "what one *calls repeating* and what one *calls dividing*") the mathematician wants to introduce back a *descriptive* component. Even if we accept Wittgenstein's introduction of a second operation within the recurring decimal's identity, we are still talking about *computational procedures*, *division* and *repetition*. And so, according to the interlocutor, we have two independent *algorithms* that fix the values of their infinite expansions quite independently of each other (contrary to what the philosopher claims). It so happens that both these *algorithms* do generate (abstractly) the same *mathematical function*.

Above and beyond any metalinguistic converse correctibility advocated by Wittgenstein, his interlocutor' retorts that this is precisely the content being asserted by that equation. One has two *abstract operations*, two *algorithms*. And the identity merely *registers a mathematical fact*, the *fact* that these two *algorithms* do describe one and the same *mathematical function*. In other words, the same intentional *object*, a function, is "in fact" denoted by both sides of that identity statement. This is what is being asserted by that statement and this is what Wittgenstein should accept. Back and forth correctibility is at best a *consequence* of that prior *mathematical fact*.

23 My italics

Wittgenstein's rejection of the notion of *function in extension*

The reintroduction of a *descriptive component* and the classical, *triangular* construal of the identity, pointing to an "abstract object", is of course precisely what Wittgenstein is advising us to *avoid*. But let us continue following the interlocutor's suggestions. Consider a computer like description of the algorithm for $[\lambda n. Div(1,7)_n]$ such as, say:

```
n = 1;
Quo_n = 1 ÷ 7;
Rem_n = 1 – (Quo_n × 7);
while (Rem_n ≠ 0)
        {
        Quo_n+1 = (10 × Rem_n) ÷ 7;
        Rem_n+1 = (10 × Rem_n) – (Quo_n × 7);
        n = n + 1;
        }
Print ("Quo_1.Quo_2 … Quo_n");
```

Why shouldn't we say that this description *fixes* the algorithm "Division of $1 \div 7$" (quite apart from any connection to any *repetition*)? After all, if we do "run it" in a computer, the digits that do come out will be "$0{,}14285714\ldots$". Why can't we say that this procedure "*determines* ahead an infinite sequence of values of that *abstract operation*", that it "*establishes*" all digits of that (recurring) expansion quite apart from any actual implementation of it by a concrete computer (or human being)? Wittgenstein writes:

> We have then a rule for dividing, expressed in algebraic or general terms,-and we have also *examples.* One feels inclined to say, "But surely the rule points into infinity-flies ahead of you – determines long before you get there what you ought to do." "Determines" – in that it leads you to do so-and-so. But this is a mythical idea of a rule –flying through the whole arithmetical series. (WITTGENSTEIN, 1976, p. 124)

> Here I should first of all like to say: your idea was that that act of meaning the order had in its own way already traversed all those steps: that when you meant it your mind as it were flew ahead and took all the steps before you physically arrived at this or that one. Thus you were inclined to use such expressions as: "The steps are *really* already taken, even before I take them in writing or orally or in thought." And it seemed as if they were in some *unique* way predetermined, anticipated--as only the act of meaning can anticipate reality. (WITTGENSTEIN, 1993, p. 64)

The philosopher refuses any idea of an *operation* "operated by *no one* in *no particular time*". Wittgenstein calls this a "mythical idea of a rule". He does accept, say, a rule connecting division and repeating, of course. And he accepts this connection as being completely independent of any empirical operations by concrete calculating agents, it is not tributary to any (empirical) reality. But he thinks it is a fundamental mistake when we move on from a atemporal and impersonal view of *rules* to an atemporal and impersonal (i.e., "abstract") view of *concepts* and *operations*. In short: Wittgenstein rejects nothing less than the very concept of a *function in extension!*

We won't be able to go into the complexities involved in Wittgenstein's *negative* to such basic classical concept. We will have to leave that for a future paper entirely dedicated to that. Besides, this would take us to far from our main objective here, the *positive* presentation of his treatment of mathematical identities. Here we will only be able to quickly sketch the main elements of that argument, as applied to our favorite example, the division $1 \div 7$. As we will see, the key element of Wittgenstein's argument against the notion of *function* can be found in Kripke's famous *Wittgenstein on Rules and Private Language* and his idea of a "*non-standard metalinguistic interpretation*" of normal ordinary functions (KRIPKE, 1972, pp. 16, note 12).

Let us go back to the idea of *back and forth correction.* When we presented Wittgenstein's proposal, we emphasized the idea that *each new rule* introduces *new possibilities of error*, new senses in which one can *disqualify* empirical claims (regarding some operations, say). As examples, we gave both the employment of *division* as correcting criteria for *repetition* (in the case when the two sequence of digits

are off by one tape's box) and *repetition* as criteria for *division* (when we discard a "result" because it contains, say, a "6" as one of the partial quotients). According to the philosopher, all these possibilities of mistake are part of the "*language mutation*" engendered by the new metalinguistic pairing of the two operations. And that much is *essential* to his conception of a "language mutation": some utterances that (before) *could describe* possible proceedings are now *discarded off hand*. We have new criteria for meaningfulness of such claims. And new possibilities of error.

With this idea in mind, let us now entertain for a while the interlocutor's *triangular view* of the equation

$$\mathrm{Div}[1,7]_{(n)} \quad = \quad 0 \frown . \frown \mathrm{Rep}["142857"]_{(\mathrm{n}-1)}$$

As we've seen, for him both sides of that identity denote one and the same "*mathematical object*", the *function in extension*

$$\left\{ (1,\mathbf{1}),(2,\mathbf{4}),(3,\mathbf{2}),(4,\mathbf{8}),(5,\mathbf{5}),(6,\mathbf{7}),\ldots \right\}$$

the "very mathematical fact being registered by that identity". So let us accept all that. Let us introduce, with him, a *third entity,* the "*extension*" (alongside with the two previous *operations*). According to our interlocutor, we would now have two "mathematical *operations*", the two "algorithms", and a "infinitary *object*", the *extension* "0,1428571428 ...".

Let us look to our situation from the point of view of Wittgenstein, now. For him, the introduction of this "*infinitary object*" would represent the acceptance of *new correcting criteria for the two previous operations*. This "mysterious object" could function now as a kind of "*Paris' standard meter*" for the two other operations. We could even introduce a new procedure: the operation of "*copying an initial section of that (infinitary) prototype*". This would then introduce a *third source of criteria* for the other two procedures. We would have *two new rules*, a rule connecting the *division* and that "*extension",* and a similar one connecting the *repetition* and the "*infinitary object".* And, as the philosopher emphasizes, the acceptance of these "new rules" would represent a *language mutation* in which it would then make

sense to talk about *new kinds of mistakes*, new criteria for the disqualification of claims regarding empirical operations.

Let us take, for example, the operation of division. Before the introduction of the "*new object*", if we wanted to claim someone was *mistaken* in his execution of the division $1 \div 7$ we would have to present him, either with a section of his calculus in which his *partial operations* were flawed, or, using the connection with the cycle "142857", point to him a section of his *result* that did not followed that pattern. With the new "*infinitary object*" though, it would suddenly make sense to accuse him of error, not because of any *internal* mistake in the division, or of an *external* mistake identified by the pattern "142857", but because his result *diverged from the (infinite) sequence of numbers stored within the mysterious infinitary standard*

0,142857142857142857142857142857142857 ...

We would now have *three* interconnected operations: the *division* $1 \div 7$, the *repetition* 142857 and the *copying* of some initial segment of that infinite standard. But, differently from the other two, this "third operation" involved an "abstract" standard *"not accessible in its entirety" by anyone, in no "empirical length of time"* (however long that may be). It would be a "guiding criteria" that would always be "partially beyond" the reach of *any agent* in *any given empirical situation*.

This is precisely where Kripke's *skeptical doubts* come in. With the introduction of this "*permanently, partially elusive object*" we have left our guards wide open to bizarre hypotheses such as those we find in (KRIPKE, 1972, p. 9). If this "Paris functional standard" makes sense, then it suddenly appears possible some kind of a strange "*undetected Babel*". What if half of the world's population has always employed one "*division's prototype*" and the other half employed a second *prototype*, but we haven't discovered the discrepancy because we haven't calculated that far? If each of us has some "function in extension", some "Division prototype" fixing ahead all decimals of *our* "infinite expansion", how do we know that our "*complete standard*" is the same as our neighbors'? What if the as yet "unsurveyed continuations" *diverge* beyond certain section?

The idea is very similar to skepticism about, say, our neighbor's "*internal sensation of green*". In both cases we have an *unreachable standard*. In the case of the *private sensation*, the standard is unreachable by anyone save the first person. In the case of the "infinite expansion object", the standard is *not ("completely") reachable by anyone*. Both constructs open ourselves up to strange forms of skepticism. Wittgenstein writes:

> The essential thing about private experience is really not that each person possesses his own exemplar, but that nobody knows whether other people also have *this* or something else. The assumption would thus be possible--though unverifiable--that one section of mankind had one sensation of red and another section another. (WITTGENSTEIN, 2005a, p. 81)

As we have said before, every new criterion introduces new possibilities of mistake. According to Wittgenstein (and Kripke), the price we would have to pay for the introduction of our "mysterious standard", i.e., the notion of *function in extension,* would be nothing less than the wild endanger of *all communicational possibilities*. Kripke's lugubrious conclusions heralds precisely this scenario:

> Of course, ultimately, if the skeptic is right, the concepts of meaning and intending one function rather than another will make no sense.
> ...if this is correct, there can of course be no fact about which function I meant. (Rules and Private language, 13)
> There can be no such a thing as meaning anything by any word. ... any present intention could be interpret so as to accord with anything we may choose to.
> What can be said on behalf of four ordinary attributions of meaningful language to ourselves and to others? Has not the incredible and self-defeating conclusion, that all language is meaningless, already been drawn? (KRIPKE, 1972, pp. 13, 54, 71)

The generalization of the idea of *back and forth correction*

Our rejection of the expression "$0{,}142\ldots$" as a *singular term* lead us to introduce a variable and a further operation of *repetition*, obtaining the *general* identity statement

$$\mathrm{Div}[1,7]_{(n)} \quad = \quad 0 \frown . \frown \mathrm{Rep}["142857"]_{(n-1)}$$

What about a *singular identity statements* such as our previous

$$2 + 3 = 5 \ ?$$

How would Wittgenstein deal with these statements?

The philosopher's final generalization of the notion of *back and forth correction* was slow to come to him. A mature, more stable view of the semantics of "*rules*" is only achieved in his final years, few years after his intermediary period represented by the *Big Typescript*.[24] But with the final explicit introduction of what we've called the "*metalinguistic component*" in his interpretation of identities, Wittgenstein is able to provide an uniform treatment of all these cases.

About the addition "$200 + 200 = 400$" he writes:

> The proof is now our model of correctly counting 200 apples and 200 apples together: that is to say, it defines a new concept: "the counting of 200 and 200 objects together". Or, as we could also say: "a new criterion for nothing's been lost or added. (WITTGENSTEIN, 1983, p. 161)

Multiplications such as "$25 \times 25 = 625$" and "$16 \times 19 = 304$" are equally interpreted as back and forth corrections:

> The fact that I have 25 x 25 nuts can be verified by my counting 625 nuts, but it can also be discovered in another way which is closer to the form of expression "25 x 25". And of course it is in the linking of

24 For a extremely careful treatment of this development, cf. (FRASCOLLA, 1994, p. Chap 2)

> these two ways *of determining* a number that one point of multiplying lies. (WITTGENSTEIN, 1983, p. 357)
>
> Multiplication could be defined by an empirical criterion. If you have 16 rows of soldiers, 19 in each row, the result by multiplication will be the same as by adding. – One feels inclined to say that if he reaches a different result from such-and-such, then he cannot mean the same by the signs as we ordinarily mean by them. "If '×' means the same, then 16×19 must have this result." (WITTGENSTEIN, 1976, p. 80)

In all these cases we have a *pairing of two different operations*. "*Whatever is taken*" as the result of the operation "counting 625 nuts" (a certain pile of nuts) is linked to "*whatever is taken*" as the result of the operation "counting 25 and 25 and then multiplying". For Wittgenstein, this "pairing" is nothing less than the "*point of multiplying*"! The metalinguistic component is clearly introduced by the philosopher: "*If* '×' *means the same, then* 16×19 *must have this result*".

Wittgenstein clearly differentiates between *empirical propositions* (involving agents and intervals of time) and (metalinguistic) *rules*. These latter do not talk about the world, about physical events, but merely lay criteria for the acceptability of such reports. Let us take the operation of "gathering 2 and 2 apples on a table". If we are talking about an *actual event* ("*in reality*"), about that *specific "gathering"* executed by *someone* at *some particular occasion*, then according to Wittgenstein we are doing an "*experiment*". If this operation is done "as we usually do it", then we could even expect to count 4 apples on the table (if this latter "counting" is done as we usually do it).

> Put two apples on a bare table, see that no one comes near them and nothing shakes the table; now put another two apples on the table; now count the apples that are there. You have made an experiment; the result of the counting is probably 4. (WITTGENSTEIN, 1983, p. 51)

This is not what rules such as "$2 + 2 = 4$" and "$4 + 1 = 5$" say. As any other rules, these do *not talk about empirical situations*, but merely lay *norms, criterions* for reports (on events involving such operations, as "*counting*" and "*adding*"):

> If we put 3 things by 2 things, that may yield various counts of things. But we see as a *norm* the procedure that 3 things and 2 things make 5 things. See, *this* is how it looks when they make 5. (WITTGENSTEIN, 1983, p. 310)
>
> It is not supposed to be an empirical proposition that the rule leads from 4 to 5, the *this*, the result, must be taken as the criterion for one's having gone by the rule. (WITTGENSTEIN, 1983, p. 319)

The number "5" should not be understood as being *defined* out of the number "4" (by the abstract "successor function"). We have a *flat pairing* of two concepts: "*counting* '4' and '1' and *gathering*" and "*counting* '5'". The first concept can function as much as correcting criteria for the second as the second towards the first.

Just as in the case of the recurring decimals, the *atemporality* and *impersonality* of the rule applies only to the "complete" statement: "2 + 3 = 5". If we take but one side of that pairing and inquire, say, "what it is to count 5?", the answer would be: "to do what one 'normally does' in those situations", what we "*normally call* '5'":

> 'If I have *five*, then I have *three* and *two*.' — But how do I know that I have five? —Well, if it looks like this: | | | | |. (WITTGENSTEIN, 1983, p. 61)

Wittgenstein's interlocutor is outraged. He misses the idea of a "mathematical operation" that *fixes ahead* the "*correct*" way of counting", quite independently of any empirical *agent* in any specific *situation*. In other words, he misses the idea of a "*function*".

Once again, we can resort to Frege on Wittgenstein' behalf, to German philosopher's idea of numerical attributions being "second order predications":

> While looking at one and the same external phenomenon, I can say with equal truth both "It is a copse" and "It is five trees", or both "Here are four companies" and "Here are 500 men". Now, what changes here from one judgment to the other is neither any individual object, nor the whole, the agglomeration of them, but rather my terminology. … This suggests as the answer … that the content of a statement of number is an assertion about a concept. (FREGE, 1978, p. 59)

Nothing *in reality* is "5". Even the dashes "| | | | |" in Wittgenstein passage could be seen as, say, an enormous number of "*ink molecules*", or an even greater number of atoms. Just as in the case of Frege's whale above, numbers do not talk (directly) about reality. Or, as Wittgenstein would say, they *can* talk about reality, but in this case there is nothing *mathematical* about them. The idea of a "*correct, atemporal and impersonal, way to count*" is just a "*mythical idea*" that should be laid aside.

Connections between Wittgenstein's ideas and contemporary Mathematics and Philosophy

We have presented, in broad brushstrokes, Wittgenstein's treatment of mathematical identities. But before we end this paper, let us quickly register what we believe are important connections between this treatment and more contemporary mathematical and philosophical literature.

The first link we would like to propose is mathematical, with the Theory of Categories of Mac Lane and Lawvere. Just as in Wittgenstein's case, the *set theoretic* concept of a *function in extension* is displaced by the *metalinguistic* concept of a *morphism*. For a category theorist, the best way to construe mathematical statements seems not so much to view them as *talking about* an ontology of (abstract) mathematical *objects*, but to step back and regard them from a metalinguistic perspective. His *morphisms* are not exactly *functions* (i.e., sets of pairs of objects), but establish metalinguistic invariants over structures. And just as in Wittgenstein's case, *identity* is somehow explained as *isomorphism* (back and forth morphisms).

The second and final connection we would like to propose is philosophical, with Quine's *Thesis of the indeterminacy of Reference* (QUINE, 1969, pp. Essays 1, 2) and his later concept of an "*observational categorical*" (QUINE, 1992, p. § 4). Right from the start, in his famous thought experiment of a linguist trying to recuperate a native's language, its lexicon and its grammar, we are invited by Quine to adopt a *metalinguistic* perspective. And, somewhat reminiscent of what happens in the theory of categories, his goal was to

distinguish contents which remain *invariant* under *radical translations* and ones that fail to do so. (QUINE, 1960, p. § 7). Once again the result of this investigation is a skepticism towards the notion of "object": according to him, this semantical notion does not seem to be "stable" under radical translations (QUINE, 1960, p. § 12). At first, Quine's attention seemed to be directed to the negative thesis, the ones about the indeterminacy of reference (and of translation), but in his final years the philosopher proposed the more positive idea of an *observational categorical*. Again what we have is a kind of semantical connection between two concepts: "whenever this, that" (QUINE, 1992, p. 10). And again this is equated with the sole content which really remains invariant over remappings:

> Translations does enjoy reasonable determinacy up through observational categorical and into logical connectives. Thus one could make a stab at the interlinguistic equating of empirical content even in radical translation. *(QUINE, 1992, p. 52)*

References

BAKER, G., & HACKER, P. (1988). *Wittgenstein: Rules, Grammar and Necessity*. Oxford: Basil Blackwell.

BRADLEY, R. (1992). *The Nature of All Being*. Oxford: Oxford University Press.

DUMMETT, M. (1977). *Elements of Intuitionism*. Oxford: Clarendon Press.

DUMMETT, M. (1991). *The Logical Basis of Methaphysics*. Cambridge: Harvard University Press.

DUMMETT, M. (1993). *The Seas of Language*. Oxford: Claredon Press.

DUMMETT, M. (1978). *Truth and Other Enigmas*. Cambridge: Harvard University Press.

DUMMETT, M. (1998). Truth from a Constructive Stanpoint. *Theoria , LXIV*, 122-38.

FRASCOLLA, P. (1994). *Wittgenstein's Philosophy of Mathematics*. Londres: Routledge.

FREGE, G. (1977a). *Logical Investigations*. Oxford: Basil Blackwell.

FREGE, G. (1964). *The Basic Laws of Arithmetic*. Berkeley: California University Press.

FREGE, G. (1978). *The Foundations of Arithmetic*. Oxford: Basil Blackwell.

FREGE, G. (1977). *Traslations from the Philosophical Writtings of Gottlob Frege*. Oxford: Basil Blackwell.

GARDINER, A. (1982). *Infinite Process: backgound to Analysis*. New York: Springer-Verlag.

KRIPKE, S. (1972). *Wittgenstein on Rules and Private Language*. Oxford: Basil Blackwell.

MARION, M. (2008). *Wittgenstein, Finitism and the Foundations of Mathematics*. Oxford: Oxford University Press.

MARTIN-LÖF, P. (1991). A Path from Logic to Metaphysics. *Atti del Congresso Nuovi problemi della Logica e della Filosofia dela Scienza. II*, pp. 141-9. Bologna: CLUEB.

MARTIN-LÖF, P. (1979). Constructive Mathematics and Computer Science. *VI International Congress for Logic, Methodology and Philosophy of Science* (pp. 1-33). Hannover: Manuscript.

MARTIN-LÖF, P. (1984). *Intuitionistic Type Theory*. Napoles: Bibliopolis.

MARTIN-LÖF, P. (1996). On the Meanings of the Logical Constants and the Justifications of the Logical Laws. *Nordic Journal of Philosophical Logic , I*, pp. 11-60.

MARTIN-LÖF, P. (1995). Truth and Kwowability: On the Principles C and K of Michael Dummett. *Second International Philosophy Conference of Mussomeli* . Typescript.

MARTIN-LÖF, P. (1987). Truth of a propostion, Evidence fo a Judgement, validity of a Proof. *Synthese , 73*, pp. 407-420.

PARSONS, C. (1983). *Mathematics into Philosophy*. Ithaca: Cornell University Press.

PRAWITZ, D. (1977). The conflict between classical and Intuitionistic logic. *Theoria* , pp. 2-39.

PRAWITZ, D. (1998). Truth and Objectivity form a verificationist point of view. In: G. OLIVERI, & H. DALES, *Truth in Mathematics* (pp. 41-51). Oxford: Claredon Press.

QUINE, W. V. (1969). Ontological Relativity. In: W. V. QUINE, *Ontological Relativity and Other Essays* (pp. 26-68). Nova Iorque: Columbia University Press.

QUINE, W. V. (1992). *Pursuit of Truth*. Cambridge: Harvard.

QUINE, W. V. (1960). *Word and Object*. Cambridge: MIT.

SHANKER, S. (1987). *Wittgenstein and the turning point in the Philosophy of Mathematics*. Londres: Croom Helm.

SKÖLEM, T. (1952,). Sobre la Naturaleza do Razonamento Matemático. *Gazeta Matemática , n° 4*, pp. 113-124.

STENLUND, S. (1990). *Language and Philosophical Problems*. London: Routledge.

VAN HEIJENOORT, J. (1967). *Form Frege to Gödel: a Source book in Mathematical Logic*. Cambridge: Harvard University Press.

WAISMANN, F. (1979). *Wittgenstein and the Viena Circle*. Oxford: Basil Blackwell.

WEYL, H. (1963). *Philosophy of Mathematics and Natural Science*. Nova Iorque: Atheneum.

WITTGENSTEIN, L. (1976). *Lectures on the Foundations of Mathematics - Cambridge 1939*. Chicago: Chicago Press.

WITTGENSTEIN, L. (2005a). *Philosophical Investigations*. Oxford: Basil Blackwell.

WITTGENSTEIN, L. (1993). *Philosophical Occasions: 1912-1951*. Cambridge: Hackett Publishing Company.

WITTGENSTEIN, L. (1975). *Philosophical Remarks*. Oxford: Basil Blackwell.

WITTGENSTEIN, L. (1983). *Remarks on the Foundations of Mathematics*. Cambridge: MIT.

WITTGENSTEIN, L. (2005). *The Big Typescript*. Oxford: Basil Blackwell.

WITTGENSTEIN, L. (1961). *Tractatus Logico-Philosophicus*. London: Routledge & Kegan Paul.

WITTGENSTEIN, L. (1976). *Wittgenstein's Lectures on the Foundations of Mathematics*. (C. Diamond, Ed.) Hassocks: The Harvester.

WITTGENSTEIN, L. (1979). *Wittgenstein's Lectures: Cambridge 1932-1935*. Totowa: Roman & Littlefield.

WITTGENSTEIN, L. (1979). *Wittgenstein's Lectures: Cambridge 1932-1935*. Totowa: Roman and Littlefield.

WRIGHT, C. (1993). *Realism, Meangin and Truth*. Oxford: Blackwell.

WRIGHT, C. (1980). *Wittgenstein on the Foundations of Mathematics*. Hampshire: Gower House.

Deferred Utterances and Proper Contexts

Marco Ruffino
Federal University of Rio de Janeiro and LanCog Group

BIBLID [0873-626X (2012) 34; pp. 807-822]

1 Proper contexts

Frege (1918-1919) famously makes a caveat concerning his doctrine of sense and reference: there are some special expressions like 'here', 'today' and 'that', that cannot by themselves express senses (and, therefore, cannot have references). Something else is necessary, namely, the extra-linguistic context in which the particular tokens of words of this kind occur. Therefore, in order to grasp the sense (and, therefore, the reference) expressed by an utterance containing some of these expressions in a particular context we need to know some facts about the latter (e.g., the place where a particular token of 'here' occurs, the day when a particular token of 'today' occurs, the demonstration that accompanies a particular token of 'that', etc.).

These expressions are what we nowadays call indexicals, and although several philosophers now reject Frege's thesis that they have a (Fregean) *sense* in a context, almost everyone retains the spirit of Frege's remark that the semantic value of indexicals (i.e., their reference or extension) depends on the context of occurrence. Moreover, almost everyone takes the semantic value of an indexical to be some element present in the context of occurrence (the day of occurrence, in the case of 'today', the utterer, in the case of 'I', the location, in the case of 'here', etc.). Different indexicals take different elements of the context of utterance (the agent, the location, the time, the addressee, etc.) as semantic value. And the semantic behavior of each indexical is governed by a rule or function that assigns a particular semantic value to each context of utterance. Following Kaplan, most philosophers call this function the *character* of the indexical, and the

semantic value assigned by the character in a particular context is called the *content* of the indexical in that context. The character is the same in all contexts, but the content might change.

Kaplan presented his theory of the semantic behavior of indexicals in his classical "Demonstratives" (Kaplan 1989), and his theory was widely accepted, to the point of becoming a kind of established view for quite a long time. One of Kaplan's thesis is that, broadly speaking, there are two sorts of indexicals: those that normally require an accompanying demonstration (typically a pointing) when placed in a context, and those that do not require such a demonstration. The first kind of expressions Kaplan calls *demonstratives*, and the second he calls *pure indexicals*. Examples of demonstratives are 'this', 'that', 'he', 'there', etc. (a demonstration is needed to make clear which of the objects in the context the speaker has in mind as the *demonstratum*); some examples of pure indexicals are 'I', 'now', 'here', 'today', etc. (no demonstration is required).

According to Kaplan's theory, pure indexicals have a special property: if they occur in a context c, they are guaranteed to have a semantic value simply in virtue of the constitutive elements of c, i.e., the speaker, the location, the time of utterance and the possible world in which c occur. This depends, of course, on what we take a context of utterance to be. In Kaplan's work we find a highly technical (and somewhat simplified) notion: a context includes an agent, a location, a time, and a possible world[1] (i.e., a world in which the utterance takes place). One important feature of Kaplan's notion of context is that its constitutive elements are always those of the production of the utterance (i.e., the agent of the context is the agent that produces[2] the utterance, the place of the context is the place of

[1] For this reason, we should probably consider 'actual' as a pure indexical as well: it has the possible world of the context of utterance as semantic value.

[2] As Perry (1997, pp. 592-3) notices, although utterances generally involve tokens, they are not always identical with the act of producing them. The agent of an utterance is not necessarily the one that produces the (written or spoken) token. In the same way, the location and time of the utterance is not necessarily the location and time in which the token is produced. The same token might be used in several different occasions (or by several different agents) to produce different utterances. For instance, someone might use a note 'Will be back soon', written by someone else, and leave it at the door of his office at different times

utterance, the time of the context is the time of the utterance, etc.)

If we think of contexts as combinations of constitutive elements (a time, a location, an agent and a possible world), we can raise the question of which of these combinations are admissible for utterances. Any combination of agents, locations, times and possible worlds is admissible? If the agent, the location, the time, and the possible world of the context are those in which the utterance is produced, then there seems to be a restriction on admissible contexts, since not just any combination of these elements are appropriate contexts for utterances. E.g., a context in which the agent is not at the location or at the time in which the utterance takes place. Thinking about these odd combinations of elements, Kaplan famously advocates their exclusion from the class of admissible contexts and a restriction to what he calls *proper contexts*:

> What has gone wrong? We have ignored the special relationship between 'I', 'here', and 'now'. Here is a proposed correction. Let the class of indices be narrowed to include only the *proper* ones-namely, those <w, x, p, t> such that in the world w, x is located at p at the time t. Such a move may have been intended originally since improper indices are like impossible worlds; no such context could exist and thus there is no interest in evaluating the extensions of expressions with respect to them. (Kaplan 1989, p. 509)

One of the efects of this restriction is that some sentences like

(1) I am here now.

come out true in any context, for in any proper context, the agent of the utterance is at the location of the utterance at the time of the utterance. (Only in improper contexts the agent might not be at the place or time of the utterance; but improper contexts are rendered irrelevant by the restriction.) A sentence that yields a true proposition in any context is a logical truth, according to Kaplan's definition.

Something interesting about (1) is that, although it is logically true, the proposition it expresses in a context is not a necessary truth. If (1) is uttered by me in Lisbon on July 9th 2010, it expresses the true proposition that M.R. is in Lisbon on July 9th 2010, but the latter is of course only contingently true. (Maybe there are other sentences

---

to make different utterances.

that are not logically true by the standard definition, but that come out as logically true if only proper contexts are allowed.) Something important in Kaplan's approach is that, by restricting the contexts to proper context, the context that is taken together with a sentence in order to produce a proposition is always the one of the production of the utterance. Therefore, the restriction to proper contexts implies that the semantic value taken by pure indexicals will be elements of the context in which the utterance occurs.

In this paper I'll be referring to this as *Kaplan's view*, but of course Kaplan is not the only one who thinks along these lines. I call it this way because Kaplan was one who clearly saw the alternatives, and explicitly and deliberately restricts contexts of utterances to proper contexts. Other influential philosophers like Montague, Lewis, Prior and Perry, to mention a few, also think that the semantic value of indexicals are taken from the context (or index) of the utterance.

Despite its intuitive appeal, Kaplan's view on this topic was challenged by some philosophers of language, based on the presentation of equally intuitive cases of utterances recorded or written in one context, but meant to be read (or decoded) in another context, i.e., at a different time, or at a different location, or uttered by a different agent. Most notably, Predelli (2005) argues that we are in many important ways free to shape the context adequate to represent an utterance, and there is no commitment either to the agent that produces the utterance, or the time or location in which the utterance takes place, or even to the possible world in which it takes place. In this paper, I shall first review these objections. Then I shall argue that these examples can be more naturally accommodated within Kaplan's theory by sharply distinguishing between a theory of semantic content from a theory of speech acts: Kaplan's approach yields an explanation of the semantic content of an utterance (which might be false) and a pragmatic theory (including, among other elements, Grice's pragmatics of implicatures) yields an explanation of how speakers can transform this semantic content into another (possibly true) content.

2 The puzzle

The phenomenon that I want to focus now, and that is sometimes regarded as an exception to Kaplan's theory, was first mentioned in print by Kaplan himself in a footnote (Kaplan 1989, p. 491, footnote 12), in which he notices that, in recorded messages for later reproduction, we find a use of 'now' that doesn't seem like a case pure indexical. E.g., the spoken message

(2) I am not here now.

recorded in an answering machine, and meant to be reproduced at different times. In cases like this, according the semantic value of 'now' doesn't seem to be the time in which the message is encoded (recorded or written), but rather the time in which the message is decoded (listened to or read).

I shall call *deferred* those written or recorded utterances employing 'now' or other pure indexicals that are meant to be read or listened to at a different time, place, etc. from that of the writing or recording. I mean deferred in a very broad sense (and not just in the temporal sense), i.e., an utterance is deferred if it is encoded in one context, but meant to be decoded in another context that differs from the original one in an aspect that is semantically relevant for some pure indexical employed in it. Deferred utterances seem to present a puzzle for Kaplan's theory: 'here' and 'now' are pure indexicals and, as such, must take as semantic values the place and time of the context in which the utterance is produced. However, this does not correspond to be the intuitive understanding of (2). How can one account for the intuitive understanding of (2) and also preserve the purely indexical nature of 'here' and 'now'?

Following a suggestion made by Donnellan, Kaplan briefly mentions in the same footnote the possibility of there being two different forms of 'now' in natural language: one that takes as reference the time when the utterance is produced, and the other that takes the time when the utterance is decoded. Let's represent both uses of 'now'. Suppose that (2) is recorded at t_0, and is decoded (i.e., heard or read) at t_1. The first use of 'now' takes t_0 as semantic value. The utterance is false (and under this reading, it cannot come out true). But the second use of 'now' takes t_1 as semantic value, and (2)

might be true (e.g., if the speaker is not at the place in which (2) is played or read in t_2).[3] However they may differ, both uses seem to have something in common, i.e., they do not need an accompanying demonstration. They automatically take some constitutive element of the context as semantic value (the time of recording or the time of decoding). So there is a sense in which they are both pure indexical uses. For brevity, I will refer to these uses as purely indexical-c (which takes the coding time as semantic value) and purely indexical-d (which takes the decoding time as semantic value), respectively.

Kaplan also mentions possible demonstrative uses of 'here' (besides the purely indexical use). One can, e.g., point at a location in a map, and say

(3) In two weeks I will be here.

I suppose that we could have something similar for 'now'. E.g., suppose that in a history lecture videotaped for future exhibitions, there is a line in the blackboard representing the chronological order of the events of the Second World War, and the teacher indicates a point of the line and says

(4) Only *now* the US enters the war.

In this use, 'now' requires an accompanying demonstration (different demonstrations of the line yielding different semantic values). This use is different from pure indexical-c and pure indexical-d above, in that simply mentioning the word 'now' won't do. A pure indexical-c reading of 'now' would have (4) saying that at the time of the videotaping the US entered the war, which is most likely false (supposing that it was videotaped long after 1941). And a pure indexical-d reading would have (4) saying that at the time in which the tape is played the US entered the war, which is also false. In this intended use of 'now', the demonstration is required.

One could think that the recognition of a demonstrative use of 'here' and 'now' would solve the puzzle, i.e., 'here' and 'now' normally work as pure indexicals, but in deferred utterances they have a demonstrative use. However, as Sidelle (1991) notices, demonstra-

[3] Multiple uses of (2) at times t2, t3, etc., presumably would pick t2, t3, etc., as semantic values.

tive occurrences of 'here' can, in principle, be replaced by 'there'. (Maybe something similar can be said for the demonstrative use of 'now', i.e., that it is replaceable by 'then'.) In the case of recorded messages, however, it does not seem right to replace (2) by

(5) I am not there now.

Therefore, what we have in the case of a recorded message is, in general, not a demonstrative use of 'here' (or now). In other words, the problem represented by utterances of (2) does not depend on the attribution of a strictly pure indexical working to 'here' and 'now'. That is to say, even if we admit of non-indexical (demonstrative) uses of 'here' and 'now', the puzzle still remains.

3 Some alternatives

There are three amendments to Kaplan's theory that I want to consider here. None of them seems to me to be adequate, for different reasons. The first amendment is suggested by Sidelle (1991) as a solution to what he calls "The Answering Machine Paradox". According to this view, there might be (and in (2) we have such a case) what Sidelle he calls *deferred utterance* (he uses the term in a different way than I do)[4], i.e., utterances in which the agent is not at the place or at the time when it occurs. Therefore, according to his view, it is not necessary for the agent of an utterance to be at the place (or time) of the utterance. According to Sidelle's proposal, when the agent records the message, he is not making an utterance at that point; he is merely arranging things so that an utterance (or maybe many of them) will be made later, at the time when someone calls the agent's machine (and the agent supposedly won't be there at the time when the utterance takes place). This seems to imply the denial of one of the principles behind the puzzle, which I will call Principle L:

(L) The utterer is always located at the time and place of the ut-

[4] It will become clearer in the last Section of this paper that Siddele's use of the term differ from mine in a subtle but important way. He means that no utterance is actually made when the utterer encodes it, but only when the listener (or reader) decodes it. I mean that an utterance (the same one) is made in both contexts.

terance.

(L) appears intuitively correct, and it is certainly part of the motivation for Kaplan's restriction of contexts to proper ones. But Sidelle thinks that it should be rejected, and he offers two reasons for this, none of them very compelling in my view. The first reason deals with (L) itself. According to Sidelle,

> [A]ll (or at least, almost all) of the sorts of utterance situations we are familiar with involve a speaker's being located at the place of his utterance. The general claim here [L] would just be an extrapolation which is perhaps not very imaginative- one, in particular, which did not focus very clearly or much on answering-machines or scraps of paper left on kitchen tables (1991, p. 534)

Sidelle's claim seems to involve some circularity; for it challenges our right to extent to messages left on answering machines and scraps of paper left on the kitchen the same account that is provided for normal utterances. This presupposes that in the latter cases, we do have something deeply different going on. But this has to be explained independently, that is to say, we need some independent reason for believing that cases like this are an exception. It is not enough to raise the vague and general doubt concerning our right to assimilate the latter cases to the former ones. Sidelle's second reason for abandoning (L) is more general, and challenges the broader claim that an agent must be located at the place where the action occurs:

> [W]e at least allow ourselves to talk in ways which don' t accord with this general claim. For instance, after putting some clothes in a washing machine, I might run into someone at the supermarket and tell them that I 'm doing my laundry (now). It would be most unusual for them to say that I can't be, since I'm not currently located by a washing machine. (1991, p. 534)

I think that examples like this bring very little support to Sidelle's claim; it seems clear to me that in the envisaged situation, 'I am doing my laundry' means that I am in middle of a somewhat complex process, which involves running to the washing machine every now and then to perform some actions that do require my presence there, and sometimes require (or allow for) my presence at the supermarket, e.g., to buy some more laundry soap. The same with 'I am finishing my PH.D. at UCLA': the fact that I am not physically on campus all the time hardly shows that I might perform an action that takes place

at UCLA without being there. The proper description of my action is that I am in a process which involves several different kinds of actions, each one of them requiring my presence, sometimes on the UCLA campus, sometimes at home (reading and writing). To sum up: in cases like those that Sidelle has in mind as counter-example to the broader thesis that underlies (L), i.e., in which the agent does not seem to be at the place where the action takes place, what most likely is going on is that the action is being misdescribed.

The second amendment is proposed, among others, by Colterjohn and MacIntosh (1987), who suggest that pure indexicals occurring in deferred utterances like (2) contain a sort of "proxy finger" indicating the time and the location in which the message is decoded. In the same spirit, Smith (1989) proposes a view according to which indexicals might have multiple characters, sometimes working as pure indexicals, sometimes working as demonstratives, and sometimes working as neither of them. For instance, the 'now' in (4) is not working according to an indexical or demonstrative rule for him, but according to what he calls the "historically emphatic" rule. (I suppose that he means by that the role of capturing a certain time that is contextually relevant, and different from the time of utterance.) Smith sees a kind of "super-rule" or what he calls a "metacharacter" governing the use of an indexical, that associates to that indexical the character appropriate to each context.

As Predelli points out (2005, pp. 48-9), there are some problems with this kind of proposal. First, regarding Colterjohn and MacIntosh's suggestion, the proxy finger is supposed to demonstrate whichever place the note is read (or listened to), but clearly many such places are not the ones that the utterer had in mind (and certainly not what the reader understands by reading the note). E.g., if (2) in a written note is carried by the wind and ends up in my neighbor's house, and if someone finds the note and reads it there he should, according to this proposal, understand that 'here' means my neighbor's house, which doesn't seem right. Second, as Predelli also correctly points out, Smith's metacharacters remain a rather mysterious semantic entities, and appealing to them looks rather like a strategy for masking the arbitrariness of characters changing from context to context. The character of an indexical is normally supposed to be a simple rule that, at least in principle, encodes once and

for all the semantic value that it should have in all different contexts. But it is hard to see anything like a general rule corresponding to the metacharacter, especially if, as a result of new devices like the answering machine, there are new characters that an indexical might have that were not foreseen.

I follow Predelli in considering the alternatives to Kaplan's view reviewed so far as unsatisfactory. For the rest of this section I will consider the amendment proposed by Predelli himself, and I will argue that it is unsatisfactory as well. His proposal is, in a way, more "conservative" than the previous ones, since it preserves the spirit of Kaplan's original approach, but at the cost of adopting a very liberal view on the notion of the intended index of an utterance. According to this proposal, indexicals always take their semantic value from an index together with which they are to be interpreted. "Index" is a technical notion, and it resembles Kaplan's notion of context: an index typically includes an agent, a location, a time, and a possible world. In Kaplan's view, as we saw, the context is always the one in which the utterance takes place, i.e., the agent, location, time and possible world in which the utterance is produced. I'll refer to the index formed by the agent, location, etc., of the utterance as the *index of utterance*. According to Predelli, whenever a speaker's utterance includes an indexical expression, a correct interpretation of the utterance must recognize that there is an intended index together with which the indexical should be taken into account, and the intended index might be different from the index of utterance. And he thinks that this explains many utterances that are intuitively true, but that would be false if taken according to Kaplan's original view. E.g., when telling someone a story involving three fictional characters A, B and C, and a fictional murder, I might say

(6) A thought all the time that B was the murderer, but *actually* C was the one who did it.

Now what is the semantic value of 'actually' in (6)? The real actual world? If so, my utterance is certainly false, for C is not a murderer in the actual world. Predelli thinks that it is natural to take the semantic value of 'actually' as being the world of the story. The same happens, e.g., in a history class about Napoleon's fate in Waterloo, when the teacher, trying to present Napoleon's situation in a dra-

matic way, speaks in the first person:

(7) What should *I* do now?

What is the semantic value of 'I' and 'now' here? If we take the index of utterance as the relevant one, the 'I' refers to the teacher, and 'now' to the time of the class. However, it is clear that the teacher was trying to say something about Napoleon in June 18th, 1815. Hence, according to Predelli, in order to correctly represent the intended utterance, we should rather take as index one that has Napoleon as agent and June 18th, 1815 as time. This is the one that the speaker had in mind in making the utterance, and the hearer can only understand the utterance if he realizes that.

In the example of the answering machine, the speaker, when recording the message, has in mind a certain index against which the message should be decoded, or, better, a set of indexes (i.e., any future instant in which the agent is not at home). The listener, in order to understand what the speaker said, has to grasp what the intended index is (supposedly, the time is the one of the phone call).

I do not find the alternative view here presented quite convincing. Most of the examples can be seen either as cases of anaphora (i.e., in the case of 'actually' in (6), it can be seen as anaphoric on the possible world described at the beginning of the report (the fictional world)), or cases in which we can plausibly see the original utterance as false but pragmatically implying a true one. (I will elaborate the latter point in the last section.) But the main obstacle for abandoning the restriction to proper contexts comes from considerations concerning the very special epistemic roles that beliefs described by pure indexicals seem to have, i.e., beliefs that one express using 'I', 'here' and 'now' (and possibly 'actual'). I can only offer here a very brief sketch of the considerations that are relevant in this connection, since a full account would require a much deeper incursion into the philosophy of mind. On the one hand, there seems to be a deep and somehow fundamental relation between the epistemic states that are expressed using pure indexicals, i.e., between those states of mind corresponding to beliefs regarding who one is, where one is, and when it is. This is, as we recall, Kaplan's main motivation for placing the restriction to proper contexts: he wanted to capture a connection between the semantic rules of 'I', 'here', 'now' (and perhaps

'actual'). The proposed revision seems to ignore, or at least to dramatically play down the relevance of this connection. On the other hand, the beliefs in question seem to play a very special role in one's total cognitive state, and are central for the believer's orientation in the world as a person and as an agent. This is part of Perry's point in calling "locating beliefs" those concerning where one is, when it is, and who one is (Perry 1979). Beliefs of this sort place the believer in a special state of mind and have motivational force that other (non-indexical) beliefs do not seem to have. Suggestions such as Predelli's, in which we are largely free to chose the context that better suits an utterance seems to miss (or at least to diminish to the point of rendering them irrelevant) the important epistemic and motivational force that these indexicals seem to have.

Maybe a better strategy, one that preserves the important epistemic roles and connections between locating beliefs would be to leave them and their contexts where they are, and look for variations in the pragmatic implications that people may draw from a proposition expressed using indexicals. There certainly is a role to be played by the speaker's intention in recording (2) or in uttering 'I will invade Russia' (in a history class). But maybe a better way of accounting for this intention is not to take the speaker's intention as playing a *semantic role*, but a *post-semantic* role[5], i.e., once every semantic value is fixed, one still has to read somehow the speaker's intention in order to draw the right conclusion from the semantically expressed proposition.

4 Back to Kaplan's original view

In this section I will sketch my proposal for dealing with deferred utterances, which holds on to Kaplan's idea that the semantic value of indexicals are always taken from the context in which the utterance is produced, and which I will call the *natural view*.[6] It is, therefore,

[5] I am using Perry's terminology (Perry 1998).

[6] Predelli calls it the "simple minded view". This designation is clearly biased, and it is hard to believe that the great founders of formal semantics would have been really simple minded at any point.

somewhat more "conservative" than Predelli's.[7]

Utterances like (2) are systematically false, but I see a divorce between what is semantically expressed by (2), and what the caller learns (i.e., a true proposition) when he hears (2). Here is a sort of consideration that, I think, supports this point of view. There seems to be a difference in our understanding of (2) recorded in an answering machine while it is being actually used as an answering machine, and when it is not being so used anymore. Imagine the situation in which there is an old recording of (2) as a greeting message left over 30 years ago by John in his old answering machine. But he has passed away some years ago, and that machine is no longer in use for answering phone calls since his widow bought a new and modern machine and recorded her own new greeting message. But the widow likes to play the old recording sometimes anyway just to hear her late husband's voice and remember the old happy days. It doesn't seem that, in this situation, the widow will take the playing of (2) as a true utterance made by the recorder, since the husband is not there anymore.

What happened? Why is it that the same message was apparently taken to be true during the many years that it was in use in the old machine(while John was alive) and now is no longer taken to be true, but is rather seen as a souvenir of John's life? (What the widow will probably think is that the recording was made at a certain place and at a certain moment in the past, and that John was certainly saying something false when he recorded (2), but which was *used for the purpose of communicating true propositions* for many years. What I think is going on here is that the semantic content of (2) is always the same, during the years of use and now, i.e., that John was not at the time and place when the recording took place, which is certainly false. But the false proposition was used (during the useful lifetime of the answering machine as an answering machine) to convey many true propositions, i.e., that John wasn't home whenever someone tried to

[7] I should remark that, as mentioned before, this proposal is not apparently the one that Kaplan himself favors in the footnote where he raises the problem (1989, p. 491, footnote 12). Although his remarks are very brief and no systematic theory for deferred utterances is elaborated in them, Kaplan seems inclined to see a pure indexical in written and recorded messages as being actually ambiguous between what I called in Section II a pure indexical-c and a pure indexical-d.

reach him during his absence. That is to say, what was decisive for taking (2) as conveying a true or a false proposition was not the semantic value of (2) itself, but the purpose for which it was employed (heard or read). In other words, what was essential was not the semantic value of (2) but, as we may call, the speech act for which it was used as a vehicle.

The same consideration could be raised concerning Smith's many character view, and in particular Colterjohn and MacIntosh's formulation of it. According to the latter authors, the note containing the indexical 'now' or 'here' works as a sort of "proxy finger" that indicates the place where the note is (or the time in which it is read). But suppose that we read the note a very long time after it was written (say, the note is exposed in a museum together with someone's writings and personal belongings). We don't take the 'here' or the 'now' in the note to refer to the museum or the present time; we see it rather as a biographical information about someone's life.

Someone thinking along the lines of Predelli's proposal could reply that, in this situation of reading or listening the message long after it was produced (and we do not take the indexical to refer to the place or time in which the message is read or listened), what happens is simply that the person who recorded or wrote it had some appropriate index (or a collection of appropriate indexes) in mind, and we are able to recognize that the present index is not one of them. But is this really so? The person who left the message in the machine seems to have in mind any future moment in which someone calls and she is not at home. Is there an intended time limit for the future indexes? The time limit is certainly a pragmatic matter; but this does not necessarily have to do with the proposition semantically expressed by the note. In order to know this proposition, we need some information regarding the situation in which it was written or recorded. But even if we do not exactly know what the original proposition was (e.g., that John was not at home at the time when the message was recorded, which is presumably false), we are able to derive, by pragmatic rules, the proposition that really matters in the present situation, i.e., that John is not home right now, at the time when we are calling him.

Something analogous can be said for another example discussed in the literature and normally presented as a counter-example to Ka-

plan's thesis. The example is:

(8) I exist.

According to Kaplan's theory, all occurrences of (8) are true because in every proper context the agent must exist, although the truth it expresses is, of course, a contingent one. This renders all occurrences of

(9) I no longer exist.

false. However, according to some critics (e.g., Predelli (2005, pp. 45-6)), true instances of (9) might occur, e.g., as part of one's will (meant to be read after the agent of the utterance has died).

Now, intuitively, if (9) is part of one's will, and if it is read short after one's death, it might be taken as conveying something true (e.g., if it is something like 'I no longer exist; since I have no living relatives, please donate all my money to charity'.). But suppose (9) appears in a will exposed in a museum of someone that died a thousand years ago. In this situation, I think it is less natural to take (9) as expressing something true. We tend to take (9) in this situation rather as a document about one's life, i.e., about what he did or thought many centuries ago. Again: what happened? If (9) expressed a true proposition right after the agent's death, it should continue to express a true proposition at any future time, but we hardly (or much less naturally) take it that way. What I think is going on is that when we look at one's will right after his death, and again many centuries later, what changes is the purpose of reading (9), and therefore the kind of proposition that we might pragmatically infer from one and the same (false) proposition expressed by (9), i.e., that that man did not exist at the time when the will was written.

If this is correct, Kaplan's original view and the corresponding restriction to proper contexts does not need to be revised. There is certainly something correct in the following complaint that Predelli formulates against Kaplan's view:

> It follows from the Simple-Minded View [...] that 'I am not here now' may never be uttered/written truly. But this clashes with our intuition that there are true instances of 'I am not here now', written on a scrap of paper or reproduced by a recording device. An analogous difficulty is raised by utterances of 'I exist (now)'. Given the thesis that 'I' and 'now' refer to the utterer and the time of utterance, together with [the

> thesis that] a speaker exists at the time of utterance (or inscription), it follows that 'I exist (now)' may not be uttered or written falsely. But this conclusion is also at odds with our intuitions concerning certain instances of written notes and recorded messages. For instance, 'I do not exist any longer' may well occur truly as part of one's will. (2005, pp. 45-6)

Predelli (and other critics) are right in that there is a clash with our intuition at some point, but the diagnosis proposed is, in my view, misleading. What we need to do is simply to recognize that the semantically expressed proposition might be taken for different purposes in different contexts, and therefore the pragmatically inferred proposition will be different. We might have the same false proposition expressed many times and in different contexts by reproducing the vehicle (note or recording) of the original utterance, but each time a different (and maybe true) proposition is inferred.[8]

Marco Ruffino
Federal University of Rio de Janeiro and LanCog Group

References

Colterjohn, J. and MacIntosh, D. 1987. Gerald Vision and Indexicals. *Analysis* **47**: 58-60.

Frege, G. 1918-1919. *Der Gedanke. Eine logische Untersuchung. Beiträge zur Philoophie des deutschen Idealismus* III, pp. 36-51.

Kaplan, D. 1989. Demonstratives. An Essay on the Semantics, Logic, Metaphysics and Epistemology of Demonstratives and Other Indexicals. In *Themes From Kaplan*, ed. by Almog, J., Perry, J., Wettstein, H. New York: Oxford University Press.

Perry, J. 1979. The Problem of The Essential Indexical. *Noûs* 13: 3-31.

Perry, J. 1997. Indexicals and Demonstratives. In *A Companion to the Philosophy of Language*, ed. by Hale, B. and Wright, C. Oxford: Blackwell.

Perry, J. 1998. Indexicals, Contexts and Unarticulated Constituents. In *Proceedings of the 1995 CSLI-Amsterdam Logic, Language and Computation Conference*. Stanford, CA: CSLI Publications.

Predelli, S. 2005. *Contexts. Meaning, Truth and the Use of Language*. Oxford: Oxford University Press.

Sidelle, A. 1991. The Answering Machine Paradox. *Canadian Journal of Philosophy* **21**: 525-39.

[8] A previous version of this paper was presented at the 1rst Annual Workshop Language, Mind and Cognition, in July 2010, in Lisbon. I thank the audience for many criticism and suggestions for improving the text. I also thank CNPq (Brazil) for the research grant that supported my work for this paper.

Ser de Uma Maneira sem Ser Claramente dessa Maneira: um Problema para o Supervalorativismo

Ricardo Santos
Universidade de Évora e LanCog Group

BIBLID [0873-626X (2012) 34; pp. 823-849]

1 A diferença entre ser F e ser claramente F e o uso do operador «claramente» para resolver o sorites

Pode uma coisa ser de uma certa maneira sem ser claramente dessa maneira? Não parece difícil encontrar exemplos de casos que sugerem que essa é uma possibilidade que, em geral, estamos dispostos a aceitar. Um predicado vago tem, ou pelo menos pode ter, casos de fronteira: objectos a respeito dos quais não é claro se o predicado se lhes aplica ou não aplica. Perante um caso de fronteira de um predicado vago «*F*», se nos for perguntado «Isto é *F*?», a reacção típica é de hesitação: não sabemos o que responder, tanto «Sim» como «Não» parecem respostas possíveis ou legítimas e, no entanto, nenhuma delas parece ter justificação suficiente. Os exemplos tradicionais são «É isto um monte?», «Este homem é careca?» ou «Esta catedral é velha?», mas podemos também pensar em casos como «Foi esta acção corajosa?», «Este indivíduo é meu amigo?» ou «Este advogado é competente?». Mas, por vezes, se a pergunta for reformulada para «É claramente *F*?», a hesitação pode desaparecer e dar lugar a uma confiante resposta negativa: «Não, *claramente F*, não é». Nesses casos, rejeitamos a descrição da coisa como «claramente *F*», ao mesmo tempo que hesitamos a respeito da sua descrição como «*F*». É plausível interpretar este género de reacção como reveladora de uma predisposição geral para admitirmos que uma coisa que não é claramente *F* pode, no entanto, ser *F*.

A admissão dessa possibilidade vai a par com a ideia de que, em geral, é preciso mais para ser claramente *F* do que para ser *F*. E que,

por isso, é natural esperar que haja menos coisas claramente *F* do que coisas *F* (numa série sorítica, por exemplo). E, se considerarmos que o advérbio «claramente» é iterável, será também natural esperarmos que haja ainda menos coisas claramente claramente *F* do que coisas claramente *F*, e assim por diante.

E, no entanto, pensadores importantes no estudo da vagueza (como Michael Dummett, Kit Fine, Rosanna Keefe, Roy Sorensen e outros) têm estado de acordo em julgarem que as coisas não podem ter uma propriedade sem terem claramente essa propriedade. Estará este seu acordo bem fundado?

Na medida em que assenta na diferença entre ser-*F* e ser-claramente-*F*, a resposta à questão que estou a colocar depende crucialmente do significado da palavra «claramente». Alguns filósofos queixam-se que esta palavra tem um significado obscuro. Mas, por outro lado, há teorias que tentam representá-la por meio de um operador frásico Δ que, num sistema formalizado, recebe uma semântica específica e para o qual são propostas regras de inferência especiais.

A queixa quanto à obscuridade do operador «claramente» pode conduzir a que se repudie o seu uso numa teorização séria a respeito da vagueza. Há, no entanto, razões aparentemente fortes para justificar a necessidade de um tal operador. Uma razão que se destaca é a necessidade que temos do operador para resolver – de uma maneira bastante *standard* na filosofia contemporânea – o paradoxo de sorites. Passemos em revista, de um modo breve, o raciocínio para esse efeito.

É comum dizer-se que o que origina o sorites é o problema da vagueza: o paradoxo explora a vagueza de certos predicados como «monte», «careca», «alto» ou «velho», mostrando que, pelo menos aparentemente, essa vagueza nos conduz a contradições. Mas em que é que consiste essa vagueza e como é que ela conduz ao paradoxo? Quer dizer: qual é a característica que os predicados envolvidos têm que podemos identificar como a sua vagueza e que seria responsável pela incoerência que o paradoxo revela? A resposta habitual a esta pergunta é a de que essa característica é a que está representada na chamada *premissa indutiva* do sorites canónico, uma premissa que diz, por exemplo, que se uma pessoa não é velha quando já passaram n segundos desde que nasceu, então ela também não é velha depois de terem passado $n+1$ segundos. Dizer que «velha» é um predicado

vago equivaleria a dizer que este princípio é verdadeiro: um só segundo não pode fazer a diferença entre uma pessoa não-velha e uma pessoa velha. Isto seria a vagueza. Quanto à incoerência, ela torna-se manifesta se observarmos que, obviamente, uma pessoa velha se tornou velha ao ter vivido um certo número, grande mas finito, de segundos: ora, não é possível que tenha havido uma transição nessa série, mas que a transição não tenha ocorrido em nenhum ponto da série (dado que a série não é mais do que uma sucessão de pontos).

Uma saída *standard* para o problema consiste então em dizer que, neste raciocínio que nos conduz da vagueza até à contradição[1], onde nos enganámos foi ao aceitarmos a premissa indutiva do argumento, concordando que um só segundo nunca pode fazer a diferença entre não-velho e velho. Embora esta afirmação nos tenha parecido ao início muitíssimo plausível, o que é facto é que o argumento mostrou que ela não pode estar certa. Como disse Crispin Wright, devemos encarar o argumento sorítico como uma *reductio ad absurdum* daquela premissa[2].

Até este ponto, muitos filósofos estão de acordo. Daqui em diante, porém, começam tipicamente a discordar a respeito de diversas questões subsequentes, como por exemplo: Devemos negar aquela premissa? Devemos considerar que ela é falsa ou devemos permanecer agnósticos acerca dela? Como se explica a forte tendência natural que temos para acreditar nela?

Menos polémica parece ser a questão de saber como é que a premissa deve ser revista, qual é o princípio que deve substituí-la. Recorde-se que começámos por aceitar a premissa porque pensávamos que ela seria a expressão adequada da vagueza do predicado. Se isso não é assim, como é então? Uma proposta bastante comum é a de dizer que aquilo a que o reconhecimento da vagueza de «velha» nos obriga é a aceitar este princípio ligeiramente modificado: *se uma pessoa* ***claramente*** não é velha quando *já passaram* n *segundos desde que nasceu, então ela também não é* ***claramente*** *velha quando já passaram* n+1

[1] A contradição é entre o reconhecimento intuitivo de que uma pessoa com 90 anos (o que é um pouco mais do que 2840 milhões de segundos) é velha e a conclusão do raciocínio sorítico segundo a qual ela não é velha.

[2] Wright (1987) 234. Ver também Fine (1975) 139 («In fact, on the supertruth view, the second premiss is false.»).

segundos. Não é entre não-velha e velha, mas sim entre *claramente-não--velha* e *claramente-velha*, que um só segundo nunca pode ser suficiente para fazer a diferença. «What the [...] paradox brings out», escreveu Wright, «is that, when dealing with vague expressions, it is essential to have the expressive resources afforded by an operator expressing *definiteness* or *determinacy*»[3].

É evidente que este género de solução do sorites coloca muito peso sobre esta noção de ser *claramente* de uma certa maneira. De acordo com esta perspectiva, o uso desta noção faz toda a diferença entre o pensamento de que a vagueza é intrinsecamente incoerente e o pensamento de que a vagueza não é paradoxal em si mesma. Um filósofo que adopte esta perspectiva encontra-se sob uma pressão muito forte para que diga algo de iluminador acerca do significado de «claramente» e acerca da diferença entre ser velho e ser claramente velho. Ele parece ser a pessoa certa a quem perguntar: pode uma coisa ser de uma maneira sem ser claramente dessa maneira?

2 Ideias centrais da teoria supervalorativista da vagueza

O supervalorativismo fornece uma interessante semântica formal para linguagens vagas, no contexto da qual a questão que coloquei pode ser analisada de um modo mais rigoroso. Tomá-la-ei como referência. Nesta secção, farei uma apresentação preliminar das principais ideias e do programa filosófico que orienta o sistema formal supervalorativista – para que, na secção seguinte, possamos analisar o seu reflexo sobre o problema que nos interessa.

Tipicamente, o supervalorativismo toma a vagueza como um fenómeno semântico, que pode afectar qualquer categoria de expressões de uma linguagem, mas que, de modo paradigmático, se manifesta nalguns dos seus predicados. Um predicado vago é um predicado cujo significado é deficiente no seguinte sentido: as suas condições de aplicação são tais que, para certos objectos, a totalidade dos fac-

[3] Wright (1987) 229. Com o operador de clareza podemos também dizer (sem ascensão semântica, na própria linguagem-objecto) que *a* é um caso de fronteira de *F*, assim: $\neg\Delta Fa \wedge \neg\Delta\neg Fa$. E podemos expressar o fenómeno da vagueza de ordem superior, dizendo que *a* é um caso de fronteira de um caso de fronteira de *F* ou, de modo equivalente, que *a* é um caso de fronteira de claramente-*F*, assim: $\neg\Delta\Delta Fa \wedge \neg\Delta\neg\Delta Fa$.

tos relevantes a seu respeito não determina se elas se encontram ou não encontram cumpridas. Face a esta indeterminação, o supervalorativismo considera que, se um tal predicado for atribuído a um desses objectos, a frase declarativa resultante poderá não ser verdadeira nem falsa. O princípio da bivalência é, por isso, rejeitado. Mas a indeterminação característica dos predicados vagos não termina aqui, nesta admissão de que, para certos objectos, o predicado nem é verdadeiro nem é falso deles. Além disso, é também indeterminado que objectos são esses. O predicado é claramente verdadeiro de alguns objectos (pertencentes à sua *extensão*) e é claramente falso de alguns outros objectos (pertencentes à sua *anti-extensão*) e não é nem uma coisa nem outra de outros objectos ainda (os casos de fronteira, pertencentes à sua *penumbra*), mas é indeterminado onde é que uns terminam e os outros começam. É por isso difícil falar da extensão de um predicado vago, pelo menos da maneira como esta noção costuma ser entendida, na teoria dos modelos clássica, como um *conjunto*. Pois não há nenhuma vagueza na noção de conjunto. A identidade de um conjunto é constituída, extensionalmente, pelos objectos que são seus membros – de tal modo que, se não estiverem definidos os objectos que lhe pertencem, não está dado nenhum conjunto. Para lidar com este problema, a ideia-chave do supervalorativismo consiste em atribuir a cada predicado vago, em vez de uma extensão definida, um leque de extensões possíveis. Estas extensões possíveis correspondem a maneiras aceitáveis de tornar o predicado preciso, ou seja, correspondem a *precisões* aceitáveis do predicado. Estas precisões são aceitáveis na medida em que são compatíveis com o significado do predicado e com os factos extra-linguísticos[4]. Nenhuma delas tem qualquer privilégio sobre as outras. Alguns objectos são tais que o predicado é verdadeiro deles numas precisões e falso noutras[5]. Mas haverá objectos dos quais o predicado é verdadeiro em todas as preci-

[4] O leitor perspicaz já terá inferido que «aceitável» é aqui um predicado vago da metalinguagem. Caso contrário, o «leque de extensões possíveis» seria um conjunto de conjuntos e estaria determinado, para cada objecto, se pertence ou não pertence a algum desses conjuntos (ou a todos ou a nenhum).

[5] Registe-se o pressuposto de que as precisões tornam o predicado *completamente* preciso, sem casos de fronteira. Ou seja, nesta versão da teoria, as precisões são interpretações clássicas, bivalentes.

sões – e aí dizemos que o predicado é *super-verdadeiro* desses objectos; de outros objectos, o predicado será *super-falso*, quer dizer, falso em todas as precisões. E a tese fundamental do supervalorativismo é a de que, para uma linguagem com predicados vagos, a noção adequada de verdade é esta noção definida de super-verdade, ou verdade em todas as precisões, entendidas como maneiras aceitáveis de tornar a linguagem precisa.

Um dos pontos fortes do supervalorativismo é a sua promessa de conciliar a existência de falhas de valor de verdade com o respeito pelas chamadas «conexões de penumbra» e pelas verdades daí resultantes. Este termo foi cunhado por Kit Fine (1975) para referir a existência de relações lógicas entre expressões vagas. Suponhamos que um certo livro tem uma cor uniforme, mas que está na fronteira entre o verde e o azul. Em certas precisões o livro conta como verde, noutras conta como azul. Mas não há nenhuma precisão aceitável em que o livro possa ser simultaneamente verde e azul. Quer dizer, a conjunção «o livro é verde e azul» é falsa, apesar de ser composta por sub-frases que são ambas indefinidas. Por outro lado, reconhecemos a disjunção «o livro é verde ou azul» como verdadeira, porque em cada maneira aceitável de precisar os predicados o livro cairá sempre na extensão de um ou na do outro. E se houver um segundo livro, situado na mesma fronteira cromática, mas ligeiramente mais azulado que o primeiro, reconheceremos como verdadeira a condicional «se o primeiro livro é azul, então o segundo livro é azul». O supervalorativismo atribui grande importância a estas verdades de penumbra. A primeira lição que retira delas é que, por causa das conexões existentes entre eles, os predicados de uma linguagem não podem ser precisados um a um, mas têm de sê-lo sempre em conjunto. Cada precisão é uma precisão da linguagem como um todo. Além disso, para ser aceitável, uma precisão tem de respeitar ou de ser fiel aos significados originais da linguagem vaga. Por exemplo, se António for um caso claro de um homem velho, nenhuma precisão de «velho» será aceitável se nela a frase «António é velho» não for verdadeira. O mesmo vale para as verdades de penumbra: elas correspondem a condições que as precisões da linguagem têm de respeitar para serem aceitáveis. Pois uma precisão em que um livro ligeiramente mais azulado do que um livro azul não fosse azul estaria a violar o significado original da palavra «azul». E o mesmo aconteceria com uma precisão

na qual um livro (todo) azul fosse também (todo) verde.

Uma outra lição que o supervalorativismo retira das verdades de penumbra é que elas mostram que, numa linguagem vaga, as conectivas proposicionais funcionam de uma maneira que é incompatível com a ideia – que é central na semântica da lógica clássica – de que elas expressam funções de verdade. Quando se abandona a bivalência e se admite que algumas frases atómicas podem ser semanticamente indefinidas, precisamos de alguma maneira alternativa para lidar com os compostos lógicos que as incluem. Para esse efeito existem propostas de novas tabelas de verdade que estendem as bem conhecidas tabelas da lógica clássica para um terceiro valor, além do verdadeiro e do falso (como, por exemplo, as chamadas tabelas fortes de Kleene). O supervalorativismo alega que as conexões de penumbra mostram que nenhuma tabela dessas pode ser correcta, pois diferentes compostos, cujos elementos têm os mesmos valores, podem receber valores diferentes. Para verificar isso, basta comparar os exemplos anteriores com estes: «o livro é verde ou interessante» (supondo que ele também é um caso de fronteira de um livro interessante) e «se o segundo livro é azul, então o primeiro livro é azul» (invertendo a ordem das frases componentes); apesar de terem a mesma forma e componentes com o mesmo valor, estas frases *não* são verdadeiras. Precisamos, portanto, de uma semântica não-verofuncional para as constantes lógicas, que seja compatível com o reconhecimento das verdades (e das falsidades) de penumbra. A solução proposta pelo supervalorativismo consiste em começar por avaliar as frases – atómicas e moleculares – em cada uma das precisões (como estas são interpretações clássicas, podemos usar as velhas tabelas bivalentes) e, depois, definir cada frase como (i) super-verdadeira, (ii) super-falsa ou (iii) indeterminada consoante ela seja (i) verdadeira em todas as precisões, (ii) falsa em todas as precisões ou (iii) verdadeira numas precisões e falsa noutras. Uma disjunção como «o livro é verde ou interessante» será verdadeira nalgumas precisões (a saber, nas precisões em que o livro caia na extensão de «verde» ou caia na extensão de «interessante») e falsa noutras (a saber, nas precisões em que o livro seja azul e desinteressante) e, por isso, é indeterminada. Já a disjunção «o livro é verde ou azul» será verdadeira em todas as precisões (pois, devido à conexão entre os dois predicados, numas precisões o livro será verde mas não azul, noutras será azul mas não verde e em

nenhuma será nem verde nem azul) e, por isso, será reconhecida pela semântica como super-verdadeira.

O supervalorativismo também fornece uma resposta interessante à questão de saber qual é a lógica correcta para os raciocínios que efectuamos numa linguagem vaga. A lógica clássica foi desenvolvida com o objectivo de formalizar o raciocínio matemático. Os seus fundadores pretendiam construir um modelo formal daquilo que, de modo informal, julgavam ser o raciocínio matemático correcto (e neste seu juízo informal enfrentaram a oposição dos intuicionistas, para quem há formas não-construtivas de raciocínio matemático que não devem ser aceites como correctas). Mas na linguagem matemática não existe vagueza (ou, pelo menos, assim parece). Por isso, não temos à partida nenhuma garantia de que a lógica clássica seja uma lógica apropriada para linguagens vagas. O argumento do sorites (que informalmente reconhecemos como válido) é válido na lógica clássica, mas isso por si só significa pouco. Pois ele poderá igualmente ser válido noutras lógicas, não-clássicas (por exemplo, ele também é válido na lógica intuicionista). No entanto, não devemos menosprezar as dificuldades envolvidas em abandonar a lógica clássica. Trata-se de uma lógica simples mas poderosa, que é bem conhecida, para a qual existem sistemas dedutivos (como a dedução natural) comprovadamente correctos e completos; e, como sabemos, é a lógica da matemática clássica e, também por essa via, de toda a ciência moderna (ou de quase toda). Além disso, se vamos discutir uma possível mudança de lógica, que lógica é que vamos usar nessa discussão?

Perante isto, uma teoria que prometa esclarecer o funcionamento de uma linguagem vaga e fazê-lo de uma maneira que permita preservar a lógica clássica poderá ser particularmente bem-vinda. E, de certo modo, é isso (ou quase) o que o supervalorativismo promete. A semântica que propõe não coincide, nem é compatível, com a semântica da lógica clássica – basta notar que a bivalência é rejeitada. Mas esta semântica não-clássica, se não acrescentarmos nenhuma constante lógica nova à linguagem, acaba por validar todas as regras de inferência da lógica clássica. Os modelos da semântica supervalorativista têm uma estrutura diferente dos modelos clássicos e as fórmulas da linguagem formal são avaliadas nesses modelos com um de três valores – super-verdadeiras, super-falsas ou indeterminadas. Um argumento é então definido como válido (a sua conclusão é uma

consequência semântica das premissas) quando é preservador da super-verdade em todos os modelos supervalorativistas. Acontece que esta noção de validade (e a relação de consequência que lhe subjaz) coincide em extensão com a noção de validade da lógica clássica: são exactamente os mesmos argumentos que são declarados válidos por uma e por outra (e as respectivas classes de verdades lógicas também coincidem). Isto é assim porque a super-verdade (aquilo que um argumento válido tem de preservar) foi definida como verdade em todas as precisões; e as precisões que compõem um modelo supervalorativista mais não são do que modelos clássicos (com o mesmo domínio). Um argumento preserva a super-verdade nos modelos supervalorativistas se e somente se ele preserva a verdade nos modelos clássicos.

Esta conservação da lógica clássica tem algumas consequências importantes com que o supervalorativismo tem de lidar. Uma delas é forçar-nos a afirmar a negação da premissa indutiva do argumento sorítico (uma vez que esta premissa, juntamente com uma outra que é totalmente insuspeita, conduz a uma conclusão manifestamente falsa). Ora, na lógica clássica, essa negação é equivalente à seguinte frase existencial: *Existe um número* n *tal que António não é velho com* n *segundos de idade e António é velho com* n+1 *segundos de idade*. Ao afirmarmos isto parece que estamos a afirmar que há, na vida de António, um segundo muito especial, com a passagem do qual ele muda subitamente de não-velho para velho. O filósofo supervalorativista tem de dar-nos uma justificação satisfatória para o facto de fazer esta afirmação – e esta é uma das suas tarefas mais difíceis[6]. Várias vias de justificação têm sido tentadas. Uma delas aponta para o facto de que, embora aquela frase existencial seja verdadeira (i.e., super-verdadeira, nos termos do supervalorativista), ela não tem nenhuma instância que seja (super-)verdadeira, quer dizer, não há nenhum número *j* referido por um termo «*k*» tal que a frase «António não é velho com *k* segundos de idade e António é velho com *k*+1 segundos de idade» seja (super-)verdadeira. De facto, na semântica supervalorativista, uma afirmação existencial $\exists x\ Fx$ pode ser verdadeira sem que nenhuma das suas instâncias o seja (e, de modo análogo, uma disjunção pode

[6] Veja-se a discussão deste aspecto em Keefe (2000) 183 e ss., confrontando com Williamson (1994) 153-154.

ser verdadeira sem que nenhum dos seus membros seja verdadeiro). Isso acontece quando todas as precisões têm um objecto que satisfaz *Fx*, mas, porque esse objecto muda de umas precisões para as outras, não há nenhum que satisfaça *Fx* em todas as precisões. Alguns autores objectam que isto não é uma leitura aceitável do quantificador existencial. Quando alguém afirma que existe um objecto que é *F*, uma reacção natural é perguntarmos «Qual?»; mas, a esta pergunta, o supervalorativista responde-nos dizendo que não é nenhum em particular. Há situações em que sabemos (por via inferencial) que existe um *F*, mas somos incapazes de identificar um objecto específico que seja *F*. A explicação normal para essas situações é o carácter limitado do nosso conhecimento. Mas a incapacidade que o supervalorativista aqui nos atribui não é devida à ignorância. Não há ignorância, neste caso, porque o que ele diz é que não há aqui nenhum facto para ser conhecido. Não há nenhum objecto para o qual o significado das palavras e os factos extra-linguísticos juntos determinem que ele seja *F*. No entanto, alegadamente, em certos casos, esses mesmos significados e factos determinam que existe um *F*.

Uma outra consequência importante da conservação da lógica clássica para o supervalorativismo é o abandono do princípio descitativo da verdade. Para a sua noção privilegiada de verdade – que é a noção de super-verdade, definida como verdade em todas as precisões –, não vale o princípio que diz que devemos aceitar todas as frases com a mesma forma que esta: «A neve é branca» é *(super-)verdadeira se e somente se a neve é branca*. Pois, para o supervalorativismo, apesar de António ser um caso de fronteira de homem velho, é verdade que ele ou é velho ou não é velho. Mas, se ele é velho, segue-se pelo princípio descitativo que a frase «António é velho» é verdadeira. E se ele não é velho, segue-se pelo princípio descitativo (e pela definição de falsidade como verdade da negação) que «António é velho» é falsa. Teríamos então de concluir que «António é velho» ou é verdadeira ou é falsa, apesar de o supervalorativista pretender que ela não é nem uma coisa nem a outra. O princípio descitativo forçaria o supervalorativista a aceitar a bivalência; e, por isso, ele rejeita-o. A sua semântica está de acordo com esta rejeição. Pois, nas precisões em que António conta como velho, a condicional «Se António é velho, então a frase "António é velho" é super-verdadeira» tem antecedente verdadeira e consequente falsa (uma vez que há outras precisões em que ele conta

como não-velho); ora, uma condicional que é falsa nalgumas precisões nunca poderá ser aceite como super-verdadeira.

A conservação da lógica clássica tem, como podemos ver, custos elevados para o supervalorativista. Além disso, é algo que ele não pode manter completamente, ou até ao fim. Pois, como vimos na secção anterior, ele precisa de dotar a linguagem de meios para expressar a noção de ser claramente de uma certa maneira; e fá-lo adoptando um operador Δ, gramaticalmente análogo à negação, que se prefixa a qualquer fórmula X para gerar uma fórmula ΔX. O supervalorativista encara este operador como uma maneira de representar, na linguagem-objecto, a noção metalinguística de super-verdade[7]. Afirmar ΔFa é, para ele, uma maneira de traduzir a afirmação metalinguística de que «*Fa*» é super-verdadeira, ou de que *a* cai na extensão de «*F*» em todas as precisões, ou de que *a* é um caso claro de «*F*». Na semântica formal, o operador Δ é tratado como uma nova constante lógica, cujas condições de verdade estão de acordo com esse entendimento intuitivo do seu significado. Na linguagem original, uma fórmula X é super-verdadeira quando é verdadeira em todas as precisões. Então, na linguagem enriquecida, diremos que, em cada precisão (de um modelo), ΔX é verdadeira se e somente se X é verdadeira em todas as precisões (desse modelo).

Com estas condições de verdade, podemos observar que, para todo o modelo M, se uma fórmula ΔX é super-verdadeira em M, então X é igualmente super-verdadeira em M; e, conversamente, se X é super-verdadeira em M, então X é verdadeira em todas as precisões de M e, por isso, ΔX também é verdadeira em todas essas precisões, pelo que ΔX é super-verdadeira em M. Nesta semântica, X e ΔX implicam-se mutuamente. No entanto, apesar desta implicação recíproca, Δ não é um operador redundante e X e ΔX não são substituíveis uma pela outra *salva veritate* em todos os contextos; pois, por exemplo, $\neg\Delta X$ pode ser super-verdadeira sem que $\neg X$ o seja (isso acontece nos modelos em que X é indeterminada). Mas, então, temos que X implica ΔX, mas $\neg\Delta X$ não implica $\neg X$, contrariamente ao que diz a regra da *contraposição* da lógica clássica (*viz.*, se uma fórmula Y se infere de X e de certas premissas e suposições, então $\neg X$ pode

[7] Como se trata de um operador, e não de um predicado, ele não permite gerar frases mentirosas (que afirmam a sua própria não-verdade).

inferir-se de $\neg Y$ e dessas premissas e suposições). Além disso, apesar de X implicar ΔX, há modelos nos quais a condicional $X \rightarrow \Delta X$ não é super-verdadeira. Mas uma das regras da lógica clássica (*a introdução da condicional*) diz precisamente que, se inferimos Y a partir de X e de certas premissas e suposições, podemos inferir a condicional ($X \rightarrow Y$) a partir dessas premissas e suposições. Isto mostra que, com a adição do operador Δ, a semântica supervalorativista deixou de conservar a lógica clássica: há novas relações de consequência (pois as inferências de ΔX para X e vice-versa não são classicamente válidas) e há relações de consequência clássica que se perdem (em argumentos que envolvem o uso do novo operador).

Além da contraposição e da introdução da condicional, há mais duas regras fundamentais da lógica clássica que falham nesta semântica supervalorativista enriquecida com o operador de clareza: a *eliminação da disjunção* e a *introdução da negação* (ou *reductio ad absurdum*)[8]. Como contra-exemplo da primeira, temos que ($\Delta X \vee \Delta\neg X$) se segue de X e também se segue de $\neg X$, mas não se segue de ($X \vee \neg X$). Como contra-exemplo da segunda, temos que, apesar de ($X \wedge \neg\Delta X$) ter como consequência a contradição ($\Delta X \wedge \neg\Delta X$), a negação $\neg(X \wedge \neg\Delta X)$ (que é equivalente a $X \rightarrow \Delta X$) não é super-verdadeira em todos os modelos.

Estes contra-exemplos de regras dedutivas clássicas que são centrais nos nossos hábitos de raciocínio devem fazer-nos parar para pensar. É verdade que, à luz da semântica supervalorativista, estas formas de raciocínio não são válidas. Mas estará o problema nas formas de raciocínio ou estará ele antes na própria semântica formal apresentada? Quando estamos a avaliar o supervalorativismo enquanto teoria da vagueza, não pode ser a sua semântica formal a ditar quais são os raciocínios que contam como correctos. A nossa avaliação intuitiva de raciocínios com termos vagos precisa de ser justificada independentemente do sistema formal. A este nível, não há substituto para a discussão filosófica.

[8] Cf. Williamson (1994) 151-2.

3 A tese supervalorativista da incompatibilidade entre X e ¬ΔX em discussão

Depois desta breve revisão dos principais traços da semântica supervalorativista para uma linguagem de primeira ordem com predicados vagos, e da sua lógica, regresso à minha questão principal de saber se uma coisa pode ser de uma maneira sem ser claramente dessa maneira – e à resposta negativa que a teoria supervalorativista *standard* lhe dá. De facto, de acordo com esta teoria, para toda a frase declarativa *X*, seria incoerente afirmar ao mesmo tempo «*X*» e «Não é o caso que claramente *X*». Por outras palavras, o supervalorativista julga que, por razões lógicas, nenhuma frase que tenha a forma «*X* e não é o caso que claramente *X*» (como, por exemplo, a frase «António é velho, mas não é claramente velho») pode ser verdadeira. Poderemos chamar a isto a *tese da incompatibilidade*[9]. E agora a questão é: por que razão deveríamos aceitar esta tese?

A tese da incompatibilidade está bem reflectida na semântica formal supervalorativista. Pois, suponhamos (para *reductio*) que há um modelo *M* no qual $(X \wedge \neg\Delta X)$ é super-verdadeira. Então, $(X \wedge \neg\Delta X)$ é verdadeira em todas as precisões de *M*. Pelas regras da conjunção e da negação, isso implica que (*i*) *X* é verdadeira em todas as precisões de *M* e que (*ii*) ΔX é falsa em todas as precisões de *M*. Mas (*i*) satisfaz as condições de verdade de ΔX (quer dizer, se *X* é verdadeira em todas as precisões, então ΔX é também verdadeira em todas as precisões), contradizendo (*ii*). Concluímos então que não há nenhum modelo no qual $(X \wedge \neg\Delta X)$ seja super-verdadeira. Esta conformidade da semântica com a tese da incompatibilidade é o que seria de esperar. Neste contexto, ela não serve de justificação para a tese, pois o que queremos avaliar é se esta semântica formal fornece um modelo adequado de como funciona uma linguagem vaga.

Uma coisa que o supervalorativista poderia estar tentado a apresentar como razão para rejeitarmos a possibilidade de *X* e $\neg\Delta X$ serem

[9] Wright (2010) chama-lhe "o Princípio de Dummett": "We can call *Dummett's Principle* the thesis that there are no truthful instances of the conjunctive form: P but not definitely P. As will emerge later, there is actually considerable pressure against the principle" (524n3). A possibilidade de rejeitar este princípio é considerada na p. 540.

conjuntamente verdadeiras é que, de $(X \wedge \neg\Delta X)$, segue-se uma contradição[10]: adoptando tal conjunção como premissa, podemos derivar ΔX e também $\neg\Delta X$. Mas é evidente que ele deriva esta contradição usando uma regra de introdução do operador Δ que o autoriza, a partir de X (apoiada, talvez, em certas premissas e suposições), a derivar ΔX (apoiada nessas mesmas premissas e suposições). Mas aceitar esta regra equivale a julgar que não é possível que X seja verdadeira e ΔX não o seja – e isto implica aceitar a tese da incompatibilidade. A regra e a tese são solidárias e não podem justificar-se uma à outra.

Haverá razões independentes para aceitar esta regra de introdução do operador Δ? Kit Fine (2008: 120) tenta fornecer uma justificação para ela. Primeiro, ele estabelece uma conexão entre as noções de incompatibilidade e de compromisso (entendidas ambas num «sentido conceptual amplo»): *a asserção de diversas proposições é conjuntamente incompatível se e somente se essa asserção nos compromete com uma contradição*. Em seguida, coloca a questão de saber se esta relação de compromisso está de acordo com a regra de introdução de Δ:

> «[...] in asserting (or in being prepared to assert) a proposition P, am I thereby committed to its being definitely the case? Surely I am. For the relevant notion of definiteness is one in which it is cognate with the notion of a borderline case. To say that x is definitely F in the relevant sense is to say that it is F and not a borderline case of F. But now the assertion that a man is bald, let us say, will surely commit one to his not being a borderline case of a bald man. For how could one sensibly assert that a given man is bald and yet not thereby be willing to deny that he is a borderline case of a bald man? Given that this is so, it will then follow directly from the above equivalence that the man is definitely bald; and the rule of D-Introduction will have been vindicated.»

A via mais habitual é explicarmos a noção de caso de fronteira em termos da noção de clareza, dizendo que *x* é um caso de fronteira de *F* se e somente se *x* nem é claramente *F* nem é claramente não-*F*. Neste trecho, Fine apresenta as coisas pela ordem inversa: toma a noção de caso de fronteira como previamente compreendida e usa-a

[10] Williamson (1994) 152: «the inference from '*p* and not definitely *p*' to 'Definitely *p*' is globally valid, as is that from the same premise to 'Not definitely *p*'». Fine (2008) 113: «The assertion that a given proposition is not definitely the case is […] presumably incompatible with asserting that it is the case, since their joint assertion would commit one to the contradiction that it is definitely the case and not definitely the case.»

para fornecer uma análise de *ser claramente F* como consistindo em *ser F e não um caso de fronteira de F*. Até aqui, não há nenhum problema. O problema começa quando, em seguida, a esta análise, Fine acrescenta uma tese substantiva, segundo a qual, de «*x* é *F*», se segue que «*x* não é um caso de fronteira de *F*». É evidente que as duas coisas juntas – a análise e a tese – são suficientes para justificar a regra de introdução de Δ. Pois, se «*x* é *F*» implica «*x* não é um caso de fronteira de *F*», então «*x* é *F*» também implica «*x* é *F* e *x* não é um caso de fronteira de *F*»; e, pela análise, esta última equivale a «*x* é claramente *F*».

Mas, nesta tentativa de justificação da regra, é notório que Fine não nos dá nenhuma razão para apoiar a sua tese de que «*x* é *F*» tem como consequência «*x* não é um caso de fronteira de *F*». A relação desta tese com a tese da incompatibilidade é muito estreita. Pois se *x* fosse um caso de fronteira de *F*, $\neg\Delta Fx$ seria verdadeira. Mas, ao propor aquela tese, Fine está a dizer que a verdade de «*x* é *F*» é incompatível com a verdade de «*x* é um caso de fronteira de *F*» e, por isso, incompatível também com a verdade de $\neg\Delta Fx$. Mas Fine simplesmente supõe que isto é assim, sem apresentar justificação. Aliás, parece típico dos defensores da tese da incompatibilidade chegarem a este ponto da discussão sem argumentos, como estes exemplos testemunham:

> «... the assertion that a man is bald, let us say, will surely commit one to his not being a borderline case of a bald man. For how could one sensibly assert that a given man is bald and yet not thereby be willing to deny that he is a borderline case of a bald man? Given that this is so...» (Fine 2008: 120)

> «We may hold that no sentence can be true without being determinately true. For how can *a* be *F* without being determinately *F*? D_p and *p* will then be true in exactly the same situations.» (Keefe 2000: 27)

> «I have heard it argued that the introduction of such an operator can serve no point since there is no apparent way whereby a statement could be true without being definitely so. That is undeniable [...].» (Wright 1987: 229)[11]

Dada esta ausência de argumentos, poderíamos tentar nós argumentar contra a tese de Fine do seguinte modo. Se de «*x* é *F*» se segue «*x*

[11] Mas, como assinalámos na nota 9, Crispin Wright dispõe-se a considerar a possibilidade de negar o «inegável» em Wright (2010: 540).

não é um caso de fronteira de *F*», então de «*x* não é *F*» seguir-se-ia «*x* não é um caso de fronteira de não-*F*». Mas não há nenhuma diferença relevante entre ser um caso de fronteira de *F* e sê-lo de não-*F*: qualquer caso de fronteira a respeito de *F* é-o tanto de *F* como de não-*F*. E uma vez que o supervalorativista aceita que, para todo o *x*, *x* é *F* ou *x* não é *F*, isso implicaria que não existe realmente vagueza: para todo o *x*, *x* não é um caso de fronteira de *F* ou *x* não é um caso de fronteira de não-*F*; mas como não há diferença conceptual relevante entre estas duas coisas, isso equivaleria a dizer que, quanto à classificação das coisas como *F* ou não-*F*, *não há casos de fronteira*. A tese de Fine, juntamente com a aceitação do terceiro excluído, conduz à negação da vagueza.

Contra este argumento, o supervalorativista objectaria que, se uma mesma coisa se segue de «*x* é *F*» e de «*x* não é *F*», isso não garante que ela se siga da disjunção «*x* é *F* ou *x* não é *F*». Ou seja, ele diria que o nosso argumento não é válido, uma vez que usa a regra da eliminação da disjunção precisamente num daqueles casos em que a semântica diz que ela não é correcta[12]. Para o supervalorativista, como já vimos, há disjunções verdadeiras sem nenhum disjunto verdadeiro. Por isso, ele considera que, quando partimos de uma disjunção e consideramos o caso em que um disjunto é verdadeiro e depois o caso em que o outro disjunto é verdadeiro, não considerámos todos os casos possíveis. Por isso, o que se segue daqueles dois casos pode não se seguir da disjunção, pois pode não se seguir do «terceiro caso» (em que nenhum dos disjuntos é verdadeiro).

Julgo que devemos resistir a esta objecção. Comparemo-la com este exemplo apresentado por Shapiro (2006: 82): um pai promete aos seus filhos que, no próximo domingo, se o tempo estiver bom irão ao futebol e, se o tempo não estiver bom, irão ao cinema. Os filhos ficam encantados porque tanto gostam de uma coisa como de outra e, por isso, concluem que vão ter um domingo óptimo. Mas chega domingo, o tempo está na fronteira entre bom e não-bom, e o pai usa isso como justificação para não fazerem nada. Não deverão os filhos sentir-se enganados e protestar que o prometido não foi cumprido? Claro que sim. O facto de o tempo se encontrar indefi-

[12] Veja-se o contra-exemplo a esta regra apresentado no final da secção anterior.

nido, nem claramente bom nem claramente não bom, significa que a família pode decidir como quiser, de qualquer das duas maneiras – mas eles têm de escolher uma delas. As promessas feitas não permitem a opção de não escolher. A situação seria diferente se a segunda promessa tivesse antes sido «se estiver a chover, iremos ao cinema». Mas, do modo como foram feitas («se o tempo estiver bom, ... e, se não estiver bom, ...»), as promessas cobriram todos os casos possíveis e, por isso, os filhos têm garantido que, se elas forem cumpridas, irão fazer algo de que gostam.

Shapiro usa o exemplo para discutir a semântica de condicionais com antecedentes vagos e consequentes precisos, mas o que aqui sobretudo me interessa é considerar o que acontece quando a família, para que o prometido seja cumprido, escolhe uma das vias. Suponhamos que, apesar de o tempo estar na fronteira entre bom e não-bom, eles decidem contá-lo como bom e vão ao futebol. Esta decisão, que é permitida pelos factos meteorológicos e pelo significado da expressão «tempo bom», não elimina a indefinição do tempo que foi inicialmente observada e consensualmente aceite pelos membros da família, quer dizer, mesmo depois da decisão tomada, o tempo (se não mudar) continua a estar na fronteira entre bom e não-bom. Mas, à luz da decisão tomada, ou no contexto por ela criado, será correcto dizer que o tempo está bom (e por isso eles vão ao futebol), embora não esteja claramente bom.

A objecção supervalorativista ao nosso argumento, ao apelar para um alegado «terceiro caso» em que nenhum dos disjuntos – nem «*x* é *F*» nem «*x* não é *F*» – é verdadeiro, assemelha-se bastante à estratégia usada pelo pai para tentar escapar ao que havia prometido. Mas não existe «terceiro caso». O que há são diversas maneiras igualmente admissíveis de estabelecer a divisão entre os dois casos possíveis – o positivo e o negativo. E, quando dizemos que elas são «igualmente *admissíveis*», estamos implicitamente a reconhecer que os casos de fronteira tanto *podem* ser classificados como positivos como *podem* ser classificados como negativos – o que é diferente de dizer que eles não são positivos nem negativos. E quando um caso de fronteira é classificado (por exemplo) como positivo, ele não deixa por isso de ser um caso de fronteira. O que está mal no argumento, por conseguinte, não é o uso da eliminação da disjunção (também chamada «demonstração por casos»), mas sim a suposição de que de «*x* (não) é *F*» se

segue logicamente que *x* não é um caso de fronteira.

Mas o que é seguir-se logicamente? Poderia argumentar-se que esta suposição é aceitável se interpretarmos a relação de consequência lógica do modo como o supervalorativista o faz, quando define a validade dedutiva como preservação da *super*-verdade. Pois se, quando raciocino a partir de «*x* é *F*», o que estou a supor é que «*x* é *F*» é *super*-verdadeira, então é natural que conclua daí que *x* não é um caso de fronteira (e o mesmo se diria para «*x* não é *F*»). Mas o que isto mostra é que temos boas razões para não aceitarmos a definição supervalorativista de validade. Quando, ao raciocinarmos numa linguagem vaga, estabelecemos uma premissa ou propomos uma suposição, para procurarmos que consequências se seguem daí, nem sempre queremos saber o que teria de ser o caso se a premissa ou suposição fosse super-verdadeira. Por vezes raciocinamos a partir de frases vagas (como «O tempo está bom», «António é velho», etc.) e queremos saber o que se segue delas tal como são, e não o que se seguiria se elas fossem super-verdadeiras. Talvez devamos então manter-nos fiéis à noção clássica de validade como preservação da verdade e rejeitar a identificação supervalorativista da verdade com a super-verdade. Diríamos que a super-verdade é uma das maneiras possíveis que uma frase tem de ser verdadeira – e certamente a melhor ou a mais estrita –, mas não é a única. Há frases que são verdadeiras sem serem super-verdadeiras. Se, no caminho para o jogo de futebol, um dos membros da família afirmar que o tempo está bom, ele não estará a violar a norma que diz que só devemos afirmar o que é verdade. No entanto, naquele contexto, «O tempo está bom» não é uma frase super-verdadeira.

Poderíamos recorrer aqui à noção de frase «suficientemente verdadeira» proposta por David Lewis (1979: 244-5). Nas suas breves observações sobre vagueza, Lewis mostra-se favorável ao método das supervalorações. Se Fred é um caso de fronteira de homem careca, a frase «Fred é careca» será verdadeira ou será falsa dependendo de onde colocarmos a linha divisória entre carecas e não-carecas. Existem diversas maneiras igualmente razoáveis de traçar essa linha e, uma vez que não há nada no nosso uso da linguagem que seleccione uma delas em detrimento das outras, devemos considerá-las todas, observando então que «If a sentence is true over the entire range, true no matter how we draw the line, surely we are entitled to treat

it simply as true». E, em seguida, Lewis escreve:

> «But also we treat a sentence more or less as if it is simply true, if it is true over a large enough part of the range of delineations of its vagueness. (For short: if it is *true enough.)* If a sentence is true enough (according to our beliefs) we are willing to assert it, assent to it without qualification, file it away among our stocks of beliefs, and so forth. [...] When is a sentence true enough? Which are the 'large enough' parts of the range of delineations of its vagueness? This is itself a vague matter. More important for our present purposes, it is something that depends on context. What is true enough on one occasion is not true enough on another. The standards of precision in force are different from one conversation to another, and may change in the course of a single conversation. Austin's "France is hexagonal" is a good example of a sentence that is true enough for many contexts, but not true enough for many others. Under low standards of precision it is acceptable. Raise the standards and it loses its acceptability.»

Esta parece ser uma visão bastante fiel de como realmente usamos uma linguagem vaga. Lewis concorda que a norma da asserção não está ligada à super-verdade. Sem violar a norma, também podemos afirmar frases que são suficientemente verdadeiras (num contexto) sem serem super-verdadeiras (nesse contexto). Além disso, as observações de Lewis apontam para uma possibilidade que, para os objectivos da presente discussão, é especialmente importante: é que o uso das supervalorações talvez não tenha de transportar consigo a tese da incompatibilidade entre *ser F* e *não ser claramente F*.

É um facto que a teoria supervalorativista *standard* inclui a tese da incompatibilidade. Vimos já que o supervalorativista não costuma perder muito tempo a procurar razões que a justifiquem. Ela parece-lhe simplesmente evidente. Vimos também que podemos desenvolver algumas linhas de argumentação contra a tese da incompatibilidade e que as objecções supervalorativistas que se lhes opõem podem ser rebatidas. Para terminar esta discussão, tentaremos agora mostrar que há uma incoerência no próprio supervalorativismo, no modo como articula a aceitação da tese da incompatibilidade com a tese de que a verdade é super-verdade e com a definição de super-verdade como verdade em todas as precisões admissíveis da linguagem (enriquecida com o operador «claramente»).

Comecemos com um aspecto do supervalorativismo que foi recentemente apontado por Delia Fara. Pensemos nalguma frase com a seguinte forma lógica: $(p \wedge \neg\Delta p)$. Como sabemos, não há nenhum

modelo supervalorativista no qual esta frase seja verdadeira (isto é, super-verdadeira). Pelas mesmas razões, uma frase que tenha a forma $(\neg p \wedge \neg\Delta\neg p)$ também não será verdadeira em nenhum modelo supervalorativista. Formemos então a disjunção destas duas frases e designemo-la por «*DF*». Podemos provar que, para todo o modelo supervalorativista *M*, se *p* é indeterminada em *M*, então *DF* é (super-)verdadeira em *M*. Isto é assim, porque, uma vez que *p* e $\neg p$ são ambas indeterminadas em *M*, tanto $\neg\Delta p$ como $\neg\Delta\neg p$ são super-verdadeiras, ou seja, verdadeiras em todas as precisões de *M*[13]; então, nas precisões em que *p* é verdadeira, o primeiro disjunto de *DF* é verdadeiro e, nas precisões em que *p* é falsa, o segundo disjunto de *DF* é verdadeiro – tendo como resultado que *DF* é verdadeira em todas as precisões e, por isso, super-verdadeira em *M*. *DF* é um dos casos de disjunção verdadeira sem nenhum disjunto verdadeiro que são característicos da semântica supervalorativista. A explicação habitual para esta anomalia é que ela resulta daquilo a que Fine chama «deslocações de valor de verdade» (*truth-value shifts*): o disjunto que torna a frase verdadeira desloca-se de umas precisões para as outras de tal maneira que, em qualquer precisão, há sempre um deles que é verdadeiro, mas nenhum disjunto é verdadeiro em todas as pre-

[13] Neste ponto do raciocínio, estamos a aplicar as condições de verdade habitualmente atribuídas a Δ, as quais são análogas às que, no sistema S5 da lógica modal, se atribuem ao operador de necessidade: $\Box p$ é verdadeira num mundo possível se e somente se *p* é verdadeira em todos os mundos possíveis; e Δp é verdadeira numa precisão se e somente se *p* é verdadeira em todas as precisões. Daqui resulta que Δp terá sempre o mesmo valor em todas as precisões de cada modelo: Δp ou é verdadeira em todas as precisões (super-verdadeira) ou é falsa em todas as precisões (super-falsa), o que faz de Δ uma noção precisa e bivalente. Estas condições de verdade atribuídas a Δ parecerão inadequadas se, por causa da vagueza de ordem superior, considerarmos que «claramente velho», «claramente claramente velho», etc., também podem ter casos de fronteira – o que deveria requerer modelos em que Δp, $\Delta\Delta p$, etc., fossem indeterminadas. Para superar este problema, Williamson (1994: 158) propõe a introdução de uma relação de admissibilidade entre precisões (em que cada precisão determinaria quais as precisões que, do seu ponto de vista, são admissíveis), sujeita a certas condições. Keefe (2000: 209-11) discute e rejeita a proposta, considerando que Δ serve para captar a vagueza de primeira ordem, mas não a sua própria vagueza (para isso, precisaríamos de novos recursos expressivos). Alguns autores pensam que a chamada «vagueza de ordem superior» não é um fenómeno real. Neste trabalho, limitamos a nossa atenção à vagueza de primeira ordem.

cisões. Essa é a explicação geral. Mas, quando se aplica a *DF*, essa explicação torna-se estranha, porque, em *DF*, os disjuntos são ambos frases que, de acordo com a semântica supervalorativista, *nunca poderiam ser verdadeiras*. Como diz Fara, o supervalorativista parece pensar que «the disjunction of two impossibilities can be true» (2010: 378).

Enquanto Fara foca a sua atenção no comportamento da disjunção, o que nos interessa aqui é sobretudo apontar o fenómeno subjacente que *DF* traz à luz e que é, muito simplesmente, o seguinte: existem modelos supervalorativistas nos quais frases com a forma $(p \wedge \neg\Delta p)$, que supostamente nunca poderiam ser verdadeiras, *são verdadeiras nalgumas precisões.* Há uma incoerência no modo como o supervalorativista lida com estas «impossibilidades». A incompatibilidade entre ser de uma maneira e não ser claramente dessa maneira é uma tese filosófica básica aceite pelo supervalorativista. E ele pode alegar que a sua adesão a essa tese está suficientemente representada no sistema formal que propõe. Pois o sistema define uma classe de modelos e não existe, nessa classe, nenhum modelo no qual uma frase com a forma $(p \wedge \neg\Delta p)$ seja super-verdadeira. Mais do que isso, o supervalorativista pode alegar que a tese que sustenta não é apenas formulável na metateoria. Ela também pode ser expressa, na própria linguagem formal do sistema, como $\neg\Delta(p \wedge \neg\Delta p)$. E $\neg\Delta(p \wedge \neg\Delta p)$ é uma *verdade lógica* do sistema supervalorativista (super-verdadeira em todos os modelos). Mas, por outro lado, em forte contraste com tudo o que o supervalorativista nos acaba de dizer, também podemos observar que existem modelos supervalorativistas que têm *precisões nas quais frases com a forma* $(p \wedge \neg\Delta p)$ *são verdadeiras*. Ora, se estas frases nunca podem ser verdadeiras e se cada precisão corresponde a «uma maneira *aceitável* de tornar a linguagem precisa» (Keefe 2000: 162), nenhuma precisão deveria avaliá-las como verdadeiras.

As precisões do sistema supervalorativista são interpretações da linguagem vaga – que, como é habitual na teoria dos modelos, incluem um domínio de objectos, uma função que atribui valores às expressões não-lógicas da linguagem e uma valoração das suas frases, simples e compostas. O que significa dizer que essas interpretações têm de ser *aceitáveis* (ou legítimas, ou admissíveis)? Uma possibilidade é considerar que as interpretações aceitáveis são aquelas que são

consistentes com as regras semânticas da linguagem[14]. Por exemplo, se há uma regra semântica como «qualquer pessoa mais alta do que uma pessoa alta é também alta», não será aceitável uma interpretação que torne o predicado «alta» preciso de um modo tal que pessoas com 1,83m contem como altas, mas pessoas com 1,86m não contem como altas. De modo análogo, a regra «nenhuma pessoa alta é também baixa» também exclui como ilegítimas interpretações que não sejam consistentes com ela, ou seja, exclui interpretações que tornem precisos os predicados «alta» e «baixa» de um modo que faça sobrepor as suas extensões. Keefe (2000: 162) dá estes exemplos como ilustrações daquilo a que Fine chama «conexões de penumbra» e, efectivamente, um dos traços característicos do supervalorativismo é a ideia de que as interpretações de uma linguagem vaga têm de respeitar as conexões de penumbra nela existentes: «Penumbral truths are true in all admissible specifications», afirma Keefe (2000: 204). E, segundo Fine, «what distinguishes th[e] operation [of making more precise] from a mere change in meaning is that it preserves truth-value» (1975: 129). Se uma frase tem um valor de verdade antes de qualquer vagueza começar a ser reduzida, ela tem de manter esse valor em todas as interpretações que tornam a linguagem mais precisa: «sentences that are unproblematically true (false) before precisification should stay true (false) afterwards» (Keefe 2000: 162)[15]. O que dizer então da frase «Nenhuma pessoa é velha sem o

[14] Williamson (1994: 158): «Supervaluationists often regard admissibility as consistency with the semantic rules of the language. If the rules decide a case, then an admissible interpretation decides it in the same way; it may decide a case when they do not.» Uma vez que a consistência é uma questão de lógica, parece que este critério faria da admissibilidade uma noção precisa – e isso seria problemático em face da vagueza de ordem superior. Williamson diz que poderíamos pensar antes num critério que fosse ele próprio vago: «Admissibility might be conceived as a matter of *reasonableness*. An interpretation is reasonable if it does not license misuses of the language (from the standpoint of an ordinary understanding of it).»

[15] No entanto, Keefe admite que «conexão de penumbra» não é uma noção precisa: «there may be vagueness over whether certain sentences qualify as penumbral connections» (204-5). E, noutra passagem (a propósito da premissa indutiva do sorites), chega a advogar uma atitude reformista a respeito *do que nos parecem ser* conexões de penumbra: «we cannot straightforwardly read off cases of penumbral truths from our unreformed intuitions, for the sorites paradox induc-

ser claramente»?

Dada a sua adesão à tese da incompatibilidade, julgo que o supervalorativista deveria contar a incompatibilidade entre «António é velho» e «António não é claramente velho» como uma conexão de penumbra e contar a frase «Nenhuma pessoa é velha sem o ser claramente» como uma verdade de penumbra. Aquela incompatibilidade ocupa, na sua teoria, um lugar muito mais central do que, por exemplo, a incompatibilidade entre «António é alto» e «António é baixo» (e, como vimos, uma precisão da linguagem na qual estas frases fossem ambas verdadeiras seria considerada inaceitável). Enquanto esta última resulta da relação entre os significados (vagos) de dois predicados não-lógicos, aquela incompatibilidade resulta da relação entre os significados (precisos ou também vagos?) de expressões a que o supervalorativista reconhece o estatuto de expressões lógicas, ou seja, é uma incompatibilidade entre formas lógicas (entre «_____ é +++++» e «_____ não é claramente +++++»). Nessa medida, assemelha-se mais à incompatibilidade entre «António é alto» e «António não é alto» (e, neste caso, são as próprias regras de valoração das frases, as quais são fixas e comuns a todas as precisões de todos os modelos, que garantem que estas duas frases nunca serão ambas verdadeiras em nenhuma precisão). Mesmo que não lhe queira chamar «conexão de penumbra» (devido ao carácter *lógico* das expressões essencialmente envolvidas), o supervalorativista deveria querer que a incompatibilidade entre «António é velho» e «António não é claramente velho» fosse respeitada pelas interpretações aceitáveis da linguagem – e que a verdade de «Nenhuma pessoa é velha sem o ser claramente» fosse preservada em todas as precisões. Mas, no seu sistema, há precisões nas quais aquelas duas frases são conjuntamente verdadeiras e esta última frase é falsa. Estas precisões não podem ser excluídas impondo restrições adicionais sobre a admissibilidade das precisões, porque elas são criadas pela própria estrutura lógica do sistema, tal como foi concebido. No entanto, à luz do modo como o próprio supervalorativista compreende a relação entre «ser

---

tive premise would then count as a penumbral truth [...]. We cannot start with all sentences that are intuitively true (both atomic predications and compound sentences) and then construct the structure of specifications so as to respect all these truths by ensuring they are true on all specifications» (183n14).

velho» e «ser claramente velho», essas precisões não podem ser vistas como interpretações legítimas da linguagem, consistentes com as suas regras semânticas ou que respeitam os seus significados. Nos termos de Williamson, trata-se de interpretações que permitem usos incorrectos da linguagem.

O supervalorativista poderia tentar defender a sua posição concedendo que a incompatibilidade entre «António é velho» e «António não é claramente velho» não é respeitada por cada uma das precisões da semântica, mas sublinhando que todos os modelos (globalmente considerados) a respeitam. Pois, nos modelos em que «António é velho» é super-verdadeira ou super-falsa, a conjunção «António é velho mas não é claramente velho» é sempre super-falsa; e, nos modelos em que «António é velho» é indeterminada (os quais representam situações em que António é um caso de fronteira), aquela conjunção é também indeterminada. É verdade que, nestes últimos, para que a conjunção seja indeterminada, ela tem de ser falsa numas precisões e verdadeira noutras; por isso, tem de haver precisões nas quais ela é verdadeira – mas isso não tem qualquer significado, é apenas um artefacto do sistema, que não representa nada a respeito do uso correcto da linguagem. É neste sentido que Shapiro (2006: 69) escreve: «For the supervaluationist, completely sharp interpretations are only a technical device used to define super-truth. They need not correspond to actual or even possible uses of the predicates.» E Keefe (2000: 190) concorda: «it is the quantification over all precisifications that captures the meaning of the natural language predicates; the individual precisifications need not.» Mais do que isso, Keefe parece conceder que, tomadas individualmente, algumas precisões podem corresponder ao que seriam usos incorrectos da linguagem: «if [...] a valuation of a language must respect all its meanings, then precisifications should not be called valuations of [a] vague language. But precisifications can still contribute to determining the correct valuation without each *being* a correct valuation in this sense» (190).

Esta linha de defesa envolve um abandono do modo como antes se caracterizou a noção de *admissibilidade*, a qual é usada para definir o conjunto de precisões (interpretações admissíveis) que são elementos de um modelo supervalorativista. Pois, enquanto antes se disse que as interpretações admissíveis seriam aquelas que respeitam as verdades de penumbra (mesmo tendo em conta a vagueza desta noção), agora

já estamos a admitir que as interpretações individuais podem não as respeitar: «we can deny that meaning *is* preserved through precisification. [...] what matters is respecting truths of meaning in the supervaluationary model as a whole» (Keefe 2000: 190n18). Com este recuo instrumentalista, o método das supervalorações perde grande parte da sua atractividade. Se as precisões individuais não correspondem a possíveis usos correctos da linguagem, mas são apenas «um dispositivo técnico», algo que se passa dentro de uma «caixa preta» e que só nos deve interessar pelos resultados que produz, então, para começar, devemos deixar de usar os conceitos de *verdadeiro* e de *falso* para falar do que se passa dentro dessa caixa. E se, quantificando sobre os elementos no interior da caixa, obtivéssemos, ao nível dos modelos, representações correctas da linguagem, que respeitassem as suas «verdades de significado», gostaríamos ainda assim de saber porque é que isso acontece. Se as frases que um modelo determina como super-verdadeiras e como super-falsas coincidissem com as frases que, num contexto de uso, são realmente verdadeiras e falsas, isto é, se a super-verdade fosse um modelo adequado da verdade, deveria haver uma explicação para isso – e uma explicação iluminadora do modo como funciona uma linguagem vaga. Em todo o caso, não parece que o supervalorativista deva estar satisfeito com os resultados produzidos pelo seu sistema de precisões. Pois ele considera que uma frase como «Ninguém é velho sem o ser claramente» é verdadeira em virtude do seu significado, mas o sistema produz modelos (com casos de fronteira de pessoa velha no seu domínio) nos quais ela *não é* super-verdadeira.

4 Observações finais

Concluo que a resposta negativa dada pelo supervalorativismo à pergunta que lancei não é convincente. O supervalorativismo não tem uma boa justificação para essa resposta e vimos, pelo contrário, que a sua adesão à tese da incompatibilidade está envolvida em diversas dificuldades, que revelam aspectos da teoria que falam em seu desfavor. Não retiro daqui nenhum juízo conclusivo a respeito do valor que se deve atribuir à abordagem supervalorativista do fenómeno da vagueza (a qual me continua a parecer, nos seus traços essenciais, uma abordagem interessante). Mas julgo que temos boas razões para

favorecer antes uma resposta positiva à pergunta inicial, ou seja, para considerar que coisas que não são claramente de uma maneira podem, no entanto, em certos contextos, ser correctamente classificadas como coisas que *são* dessa maneira – e que é permissível afirmar que elas o são, sem que com isso deixe de ser verdade que elas não são claramente como as estamos a classificar. Dito de outro modo, temos boas razões para considerar que, numa linguagem vaga, há frases que, afirmadas em certos contextos, são verdadeiras, embora não sejam claramente (ou determinadamente) verdadeiras – frases que, como diz Lewis, para os padrões de precisão aceites num contexto conversacional, são *suficientemente verdadeiras*, mas não mais do que isso. É uma questão em aberto a de saber se a abordagem supervalorativista é adaptável a esta perspectiva de um modo que possa ser descrito, usando a conhecida expressão de Tarski, como «formalmente correcto e materialmente adequado»[16].

No final da sua monografia, Keefe (2000: 215) recorda as célebres palavras de Aristóteles (que Tarski costumava citar em apoio do esquema descitativo): «Dizer daquilo que é que não é, ou daquilo que não é que é, é falso; enquanto dizer daquilo que é que é, ou daquilo que não é que não é, é verdadeiro.» Keefe pensa que, ao dizer isto, Aristóteles não cobriu todos os casos possíveis e, por isso, pergunta (por exemplo) como é que classificaríamos o «dizer, daquilo que indeterminadamente é, que é». A minha resposta seria: há situações em que isso também é verdadeiro.[17]

Ricardo Santos
Universidade de Évora e LanCog Group

[16] McGee e McLaughlin (1995) e Shapiro (2006) trabalham nessa direcção.

[17] Este trabalho foi realizado no âmbito do projecto de investigação *Paradoxos: Dedutivos, Indutivos e Práticos* (PTDC/FIL/67039/2006), financiado pela Fundação para a Ciência e a Tecnologia e sediado no Instituto de Filosofia da Linguagem FCSH-UNL e na Universidade de Évora. Apresentei versões preparatórias dele no Seminário de Filosofia Analítica da Faculdade de Letras da Universidade de Lisboa, num *workshop on vagueness and self-reference* que teve lugar na Faculdade de Ciências Sociais e Humanas da Universidade Nova de Lisboa e numa palestra que proferi na Universidade Federal do Rio de Janeiro. Agradeço aos participantes nessas sessões a atenção dispensada e as estimulantes discussões que se seguiram. Agradeço em especial os comentários valiosos que recebi de João Branquinho, Rosanna Keefe, Elia Zardini, Hartry Field, Oswaldo Chateaubriand e Guido Imaguire.

Referências

Dietz, Richard e Moruzzi, Sebastiano (eds.). 2010. *Cuts and Clouds: Vagueness, Its Nature, and Its Logic*. Oxford: Oxford University Press.

Fara, Delia Graff. 2010. Scope Confusions and Unsatisfiable Disjuncts: Two Problems for Supervaluationism. In *Cuts and Clouds: Vagueness, Its Nature, and Its Logic*, ed. por Dietz e Moruzzi. Oxford: Oxford University Press, pp. 373-382.

Fine, Kit. 1975. Vagueness, Truth and Logic. In *Theories of Vagueness*, ed. por Keefe e Smith. Cambridge: Cambridge University Press, 2000, pp. 119-150.

Fine, Kit. 2008. The Impossibility of Vagueness. *Philosophical Perspectives* 22, pp. 111-136.

Keefe, Rosanna. 2000. *Theories of Vagueness*. Cambridge: Cambridge University Press.

Lewis, David. 1979. Scorekeeping in a Language Game. In *Philosophical Papers*, vol. I, Oxford: Oxford University Press, 1983, pp. 244-245.

McGee, Vann e McLaughlin, Brian. 1995. Distinctions Without a Difference. *Southern Journal of Philosophy* 33 (supplement), pp. 203–251.

Shapiro, Stewart. 2006. *Vagueness in Context*. Oxford: Clarendon Press.

Williamson, Timothy. 1994. *Vagueness*. Londres e Nova Iorque: Routledge.

Wright, Crispin. 1987. Further Reflections on the Sorites Paradox. In *Theories of Vagueness*, ed. por Keefe e Smith. Cambridge: Cambridge University Press, 2000.

Wright, Crispin. 2010. The Illusion of Higher-Order Vagueness. In *Cuts and Clouds: Vagueness, Its Nature, and Its Logic*, ed. por Dietz e Moruzzi. Oxford: Oxford University Press, pp. 523-549.

Ceteris Paribus Laws and the Human Sciences

Rui Silva
University of the Azores and LanCog Group

BIBLID [0873-626X (2012) 34; pp. 851-867]

In opposition to the traditional conception of scientific laws as universal and exceptionless generalizations that hold everywhere and throughout all times, some philosophers of science have proposed the idea of a *ceteris paribus* law (CP law), i.e., a law that is true "all else being equal", in the sense that it can admit of exceptions because of the influence of interfering factors. The legitimacy of CP laws has divided the philosophical community. Some argue that all scientific laws are CP laws, others argue that CP laws are simply unacceptable, and a third group claims that science contains both CP laws (in the so-called special sciences) and strict laws (in basic physics).

The debate on CP laws has important implications for our understanding of the human sciences, because if we accept the existence of laws in sciences like psychology or sociology, these laws, far from being strict, should be CP laws. Furthermore, if we could establish some differences between CP laws in the human and the natural sciences, such a difference could help us to understand the relation between these two scientific domains.

In what follows, I will first discuss the concept of CP law and present some proponents of such laws (1). In order to explore the significance of this debate for the human sciences, I will discuss Davidson's and Fodor's stance on the status of CP laws in the human sciences (2). After these introductory sections, I will argue that in the human sciences it is much more difficult, not only in practice, but also in principle, to cope with interfering factors and to refine *ceteris paribus* clauses (3). Against hermeneutic and strict naturalistic perspectives, I propose an account of the human sciences that rec-

ognizes the methodological role of both generalizations and rationality assumptions; in particular, I will stress the need for interpretative/hermeneutic methods to control the vagueness of *ceteris paribus* clauses (4).

1

A fundamental problem in the philosophy of the human sciences consists in the existence or possibility of laws in these sciences. Because of the complexity of the factors that can influence human behaviour, generalizations in psychology and sociology, for instance, have a somewhat fragile status, especially when they are couched in an intentional vocabulary. Predictions based on such generalizations may well fail, and it is for this reason that the idea of CP laws, understood as laws that admit of exceptions, becomes relevant for understanding the human sciences.[1]

A CP law describes real tendencies or causal powers, but it idealizes the phenomena to be explained and abstracts them from the possible interference of factors that lie beyond its scope. It is in this sense that CP laws admit of exceptions and can be contrasted with strict laws. According to the traditional conception, scientific laws are universally valid, empirically confirmable and counterfactual-supporting generalizations. They can be either strict or statistic, but in both cases they are conceived as exceptionless. A CP law, on the other hand, is a generalization that is true "all else being equal", i.e., exceptions are possible because of the interference of an indefinite number of factors that cannot be specified in advance.

The idea of CP laws is, of course, controversial. In the first place, these laws seem to be vacuous. Let us consider the following formulation of a law with a *ceteris paribus* clause: "All As are Bs, except when they are not". This is a caricature of a CP law, but it points to a real problem with CP laws, its apparent vacuity. Closely related to this objection is the charge of vagueness. CP laws are true all else being equal. But what is exactly "all else" (*ceteris*)? And what is the exact

[1] *Ceteris paribus* laws have already been labelled as soft laws (cf. Horgan/Tienson 1996), and they are, as such, appropriate candidates for the role of laws of the so-called soft sciences.

meaning of "being equal" (*paribus*) in a particular context? If CP laws lack determinate truth conditions, advocates of truth-conditional theories of meaning can doubt that they have meaning.

Last, but not least, CP laws seem to evade the requirements of testability or falsifiability. The idea that laws may have exceptions because of the interference of unexpected factors threatens to immunize them against empirical refutation; it allows the scientist to hold a law in the light of recalcitrant evidence by introducing *ad hoc* hypotheses.[2]

However, and according to their proponents, CP laws are legitimate. To begin with, there are many scientific generalizations that have exceptions but are considered as genuine laws; Boyle's law and the law of supply and demand are two good examples. The supporters of CP laws also claim that, in order to study a very complex world, science has inevitably to appeal to abstractions and idealizations. CP laws belong to these idealizations; they single out causal tendencies that, far from operating in isolation, interact with many other causal tendencies. Since events result from a combination of causes, a real tendency may be overridden by other causal forces, and this would entitle us to introduce *ceteris paribus* clauses in law-statements.

Carl Hempel's reflections on the role of provisos was a major source of inspiration for supporters of CP laws. He claimed that Newton's law of gravitation involves certain assumptions and can lead to inaccurate predictions if, for instance, non-gravitational forces interfere with gravitational ones. As a result, predictions based on Newton's law are true, *provided* that no significant external forces (gravitational or non-gravitational ones) affect a planetary system. Since an indefinite number of forces may act on the planets, the proviso seems to have an open-ended character similar to *ceteris paribus* clauses. For this reason, some authors attributed to Hempel the thesis that even the fundamental laws of physics contain *ceteris paribus*

[2] In the words of Roberts (2004: 162): "hedged regularities cannot be discovered by science, because they are not empirically testable. The hedge functions as an 'escape clause' that allows any hypothesized hedged regularity to escape empirical refutation: Whenever you discover a counterexample, claim that there has been an interference of some kind, so that the case is outside the scope of the hypothesis and does not falsify it".

clauses (cf., e.g., Fodor 1991). This interpretation of Hempel's conception of provisos is, however, dubious. Earman and Roberts (1999: 442-446) have argued, in my view persuasively, that the conditions of the Hempelian provisos are not relative to the truth of law statements, but only to the validity of the application of a theory. This point distinguishes his provisos from current accounts of *ceteris paribus* clauses.

Another important voice in the debate on CP laws is Nancy Cartwright. She was known for a long time as a supporter of CP laws because of her opposition to a Humean understanding of laws as regularities and her thesis that scientific laws should be understood in the "language of *powers*, *capacities* or *natures* and related concepts such as *interfere*, *inhibit*, *facilitate*, and *trigger*" (Cartwright 2002: 150). She has, indeed, claimed that scientific laws are CP laws in the sense that "they hold only relative to the successful repeated operation of a nomological machine" (Cartwright 1997: 66). By nomological machine she understands a set of capacities that cause in a stable environment the regularities that are described by scientific laws. Nomological machines are also described in her work as the object of scientific models.[3]

Fodor (1991) and Pietroski/Rey (1995) represent another influential defence of CP laws. In spite of some differences, their accounts of CP laws share some common assumptions, namely that scientific explanations rely on laws and that strict laws must be exceptionless. Since many scientific generalizations admit of exceptions, they claim that for a given CP law there is a condition that, when realized, eliminates exceptions. Fodor calls such a condition a "completer". The combination of a CP law with the corresponding completer amounts to a strict or exceptionless law. In the case of Pietrosky and Rey, CP laws are considered genuine laws if prediction failures can be explained by a fact that has an independent explanatory role. More precisely, an exception to a scientific law is legitimate if we can indicate an interfering factor and if this factor can be explained on the

[3] More recently, Cartwright has distanced herself from CP laws, but the point is apparently terminological; she does not agree with characterizations of CP laws as laws that neither can be stated in precise and closed form nor entail strict or statistical regularities (cf. Cartwright 2002: 149). At any rate, she remains faithful to an understanding of laws in terms of powers and capacities.

basis of an independent theory. This condition allows us to prevent *ad hoc* manoeuvres in science. Pietroski and Rey illustrate this point in the following metaphorical terms:

> But there need be no intrinsic problem about ineliminable CP-clauses. Briefly, we claim that such clauses are cheques written on the bank of independent theories, their substance and warrant deriving from the substance and warrant of those theories, which determine whether the cheque can be cashed." (Pietroski/Rey 1995: 82)

This approach has the advantage of defending CP laws from the charge of vacuity without trying to give determinate truth conditions to them.

At a more methodological level, there are some elaborate attempts to establish the legitimacy of CP laws. A case in point is Harold Kincaid. In chapter 3 of his *Philosophical Foundations of the Social Sciences*, he claims that, appearances notwithstanding, CP laws are serious and testable laws. It would be inappropriate in this context to delve into his reflections on the testability problem, but one can highlight some of his main points. He stresses, for instance, that real situations fit sometimes scientific idealizations, and in these cases CP laws become clearly testable. Furthermore, exceptions can be often explained away by reference to well-established laws or causal mechanisms, and successive refinements may increase the accuracy of the predictions. Kincaid also stresses that CP laws may predict novel facts and support counterfactuals.

He cites in this context the work of the sociologist Jeffrey Paige on the relations between agrarian structure and political behaviour as an example of good social science that offers us lawlike generalizations (Kincaid 1996: 70-80). By gathering the relevant evidence, Paige formulated a set of interesting generalizations that do not express strict regularities, but have other traits of the traditional conception of law; they are, namely, empirically confirmable and counterfactual-supporting. Here are some examples of these generalizations:

1. Plantation systems (where owners depend on capital and workers on wages) encourage collective action and labor reforms.
2. Commercial hacienda systems (where both owners and cul-

tivators get their income from land) favour agrarian revolts.
3. Large estate systems (where owners get their income from the land and cultivators from wages) tend to lead to revolutionary movements.

Paige was able to confront these generalizations with the available data and concluded that they hold generally. The generalizations include, of course, *ceteris paribus* clauses, but one of the merits of his work consists precisely in showing how a researcher can refine the *ceteris paribus* clauses by considering the relevant interfering factors and by explaining their influence on the causal tendencies expressed by the generalizations. For instance, he shows how *contagion effects* (the increased probability of an event on the basis of previous occurrences of similar events) and the action of urban political parties influence the political behaviour of cultivators.

What can we conclude from the preceding considerations? It would be too ambitious to try to settle here the intricate debate on the legitimacy of CP laws.[4] For the purposes of the present article, it is sufficient to clarify the concept of CP law, to understand the motivation behind it, to evaluate the significance of scientific generalizations with *ceteris paribus* clauses and to compare their status in the natural and in the human sciences. In fact, even if one refuses to accept CP laws as genuine laws, one may well accept that there are many scientific generalizations that admit of exceptions, but are testable and important for science.[5]

As we have seen, many respectable lawlike statements must inevitably idealize phenomena that, in the real world, are the result of very complex causal interactions that cannot be grasped by unqualified universal statements. Lawlike generalizations typically describe

[4] For a good survey on the main accounts of CP laws and their difficulties, see Earman/Roberts 1999.

[5] A case in point is Jim Woodward, who denies the existence of CP laws, but accepts the corresponding generalizations: "While I reject the idea that generalizations found in the special sciences are ceteris paribus laws, I fully agree that many of those generalizations are 'scientifically legitimate', that they are testable and in fact strongly supported by evidence [...]. My claim is that construing those generalizations as ceteris paribus laws is the wrong way to defend their usefulness and legitimacy" (Woodward 2002: 306).

a causal tendency, but a tendency that interacts in the real world with other tendencies. This lends to generalizations with *ceteris paribus* clauses a prima facie legitimacy. However, we should keep in mind that the existence of *ceteris paribus* clauses weakens a nomological statement and may undermine its explanatory power.

2

In order to evaluate the significance of CP laws for the human sciences it is convenient to consider the work of Davidson and Fodor in this context. Both of them accept the existence of CP laws in the special sciences, but evaluate differently their role in the human sciences.

Davidson argued in a series of influential papers that there are no strict psychological laws. By strict law he understands the traditional conception of law as a non-accidental generalization, characterized by being empirically confirmable and by supporting counterfactuals. Intentional psychology produces many useful generalizations, but, according to Davidson, they are not strict laws; far from being exceptionless, they are vulnerable to the interference of an indefinite number of factors that cannot be specified in advance.

His defence of the Principle of the Anomalism of the Mental is based on a rejection of psychophysical laws. The basic point behind psychophysical anomalism is the claim that the vocabularies of physics and psychology are incommensurable:

> There are no strict psychophysical laws because of the disparate commitments of the mental and physical schemes. It is a feature of physical reality that physical change can be explained by laws that connect it with other changes and conditions physically described. It is a feature of the mental that the attribution of mental phenomena must be responsible to the background of reasons, beliefs, and intentions of the individual. (Davidson 1980: 222)

According to Davidson, the gap between the vocabularies of physics and psychology results from the holistic and normative character of the psychological domain. The content of mental states is partly determined by their reciprocal relations and its interpretation is based on rationality assumptions:

> Any effort at increasing the accuracy and power of a theory of behaviour forces us to bring more and more of the whole system of the agent's beliefs and motives directly into account. But in inferring this

> system from the evidence, we necessarily impose conditions of coherence, rationality, and consistency. These conditions have no echo in physical theory, which is why we can look for no more than rough correlations between psychological and physical phenomena" (Davidson 1980: 231)

Another key point of the Davidsonian defence of the anomalism of the mental departs from the idea that a strict or exceptionless law requires a closed domain, like the domain of physics. In order to explain physical phenomena we do not need to accommodate the possible interference of non-physical causes; on the contrary, physicists and philosophers of science endorse the principle of the causal closure of physics. But the psychological domain is clearly not closed. Psychological phenomena are exposed to the action of non-psychological causes; biological, chemical and physical factors. For this reason, we cannot formulate a psychological theory that provides a complete coverage of the psychological domain. Psychological anomalism, the thesis that there are no strict laws describing correlations between psychological phenomena, can be presented as a consequence of the conjunction of psychophysical anomalism and the claim that the psychological domain is not closed (cf. Davidson 1980: 224).

Davidson is aware that scientific laws may admit exceptions, but he claims that in the natural sciences it is possible, in principle, to refine and improve laws in order to eliminate exceptions. Psychological generalizations are not refinable in this sense. Thus, in "Psychology as Philosophy" he clearly dismisses the possibility of strict laws dealing with beliefs and desires, the kind of laws that could ground intentional explanations of the human behaviour. In his perspective, it is hopeless to expect laws of the type "whenever a man has such-and-such beliefs and desires and such-and-such further conditions are satisfied, he will act in such-and such a way" (Davidson 1980: 233). He claims that:

> What is needed in the case of action, if we are to predict on the basis of beliefs and desires, is a quantitative calculus that brings all relevant beliefs and desires into the picture. There is no hope of refining the simple pattern of explanation on the basis of reasons into such a calculus. (Davidson 1980: 233)

The lack of precision of psychological laws cannot be explained in the same way that we explain the lack of precision of some natural

scientific laws; the laws involved in meteorological explanations and forecasts, for instance. The problem with psychological generalizations does not consist simply in the complexity of the factors that intervene in the human behaviour. Meteorological forecasts are admittedly fallible because of the complexity of meteorological phenomena, and earthquakes, according to the current status of seismology, cannot be predicted. Serious laws may have *ceteris paribus* clauses in order to accommodate the possible interference of factors that generate exceptions to them, but Davidson thinks that there is, at any rate, an important distinction between the *ceteris paribus* clauses of natural scientific laws and the corresponding clauses in putative psychological laws. In the former case, but not in the latter, it is possible, in principle, "to determine in advance whether or not the conditions of application are satisfied" (Davidson 1980: 233). This claim should be understood in the light of Davidson's holistic and normative conception of the psychological domain. Because different individuals have different sets of beliefs and desires, generalizations couched in an intentional vocabulary express mere tendencies and cannot aspire to the degree of precision that characterizes natural scientific laws. Furthermore, any attempt to explain and predict actions on the basis of beliefs and desires must rely, according to Davidson's Principle of Charity, on rationality assumptions, and these assumptions are not rigid; there are different ways to rationalize an agent's behaviour.

Davidson's defence of psychological anomalism is based on important insights, but is not satisfactorily developed in his work. In particular, he does not clarify the key notion of *ceteris paribus* law, although he appeals to laws of this kind, namely when he argues for the causal efficacy of mental events. In "Thinking Causes", Davidson appeals to non-strict laws in order to save his anomalous monism from the charge of epiphenomenalism. In fact, according to Davidson's principle that causal relations are covered by laws, he seems to be forced to appeal to psychophysical laws if he wants to attribute causal powers to mental states, but since strict laws cannot be couched in a mental vocabulary, laws that correlate mental events with physical events should be non-strict laws. A further advantage of these laws consists in the fact that they do not entail a physical reduction of the

mental. [6] Non-strict laws have, according to Davidson, the merit of explaining the causal efficacy of the mental while avoiding at the same time the threat of reductionism.

Fodor claimed, in opposition to Davidson, that the status of psychological generalizations does not differ from the status of many other laws in the natural sciences. Fodor concedes that intentional laws are non-strict laws, but does not think that this point raises a particular epistemological problem for psychology, because sciences like biology, meteorology or geology rely also on non-strict laws. Nomological generalizations in the so-called special sciences admit of exceptions, because of the presence of interfering factors that cannot be taken into account in the formulation of the law, but it is usually possible to explain exceptions to a law in the vocabulary of another, more basic sciences. In the case of psychology, exceptions to generalizations could be explained with the resources of sciences like neurology or biochemistry:

> Exceptions to the generalizations of a special science are typically *inexplicable* from the point of view of (that is, in the vocabulary of) that science. That's one of the things that make it a *special* science. But, of course, it may it may nevertheless be perfectly possible to explain the exceptions in the vocabulary of some other science. (Fodor 1987: 6)

Fodor argues that the same holds for commonsense psychology; the *ceteris paribus* clauses of its generalizations are "ineliminable from the point of view of its proprietary conceptual resources", but "can be discharged in the vocabulary of some lower-level science" (Fodor 1987: 6).

3

Against Fodor, I argue that there are significant differences between psychological generalizations and laws of natural science. CP laws are *prima facie* legitimate because of the existence of interference effects

[6] Davidson was sensitive, in his context, to Kim's following objection: "The trouble is that once we begin talking about correlations and dependencies between specific psychological and physical properties, we are in effect talking about psycho-physical laws, and these laws raise the spectre of unwanted physical reductionism. Where there are psycho-physical laws, there is always the threat, or promise, of psycho-physical reduction." (Kim 1993: 278-9).

between different causal factors, but we must distinguish between two different types of interferences. The first type is what I will call *external interference*, and it is present whenever an exception to a generalization is caused by factors that belong to another domain, which is covered by another theory or science. It is dubious that this kind of interference has the same form in psychology and in natural sciences like biology or geology. In fact, psychological generalizations cannot be linked to, say, neurology in the same way that biological processes, for instance, can be linked to chemical processes. Psychological entities are partly constituted, as Davidson argued, on the basis of rationality assumptions and normative principles, and this sets them apart from the world of natural phenomena. By studying the chemical basis of a biological phenomenon one may understand it better, but neurological studies do not contribute to our understanding of reasons. McDowell's distinction between the "space of reasons" and the "realm of laws" may clarify this point. Those who argue that the relation between psychology and the corresponding lower-level theories (such as biology or chemistry) is analogous to the relation between higher-level and lower-level theories in the natural sciences neglect the fact that in the former case we are in presence of two different forms of intelligibility; in the latter case, we are dealing with different levels in the same domain (the "realm of law"). This means that in psychology and other human sciences the phenomena of external interference are much more difficult to handle (not only in practice, but also in principle).

The second form of interference, *internal interference*, occurs when the interfering factors belong to the theoretical domain of the generalization. In this regard, there is an important difference between psychology (as well as other human sciences) and the natural sciences. In the latter sciences, there are no in-principle obstacles to a gradual refinement of *ceteris paribus* clauses and to a formulation of laws that are exceptionless in their own domain, but there are good reasons to think that the same does not hold for laws involving intentional notions. Any generalization in terms of beliefs and desires is always threatened by the interference of competing beliefs and desires in a way that challenges any theoretical approach. Two arguments can be invoked in this context. In the first place, and assuming as true a holistic account of the psychological realm, the fact

that the content of mental states is determined by their location in a network of propositional attitudes and that different interpretations may determine differently the content of beliefs, desires or intentions makes it impossible to fix all the variables involved in intentional explanations. In the second place, John McDowell's reflections on the uncodifiability of human rationality and, in particular, his thesis that human actions cannot be explained in terms of deductions from a definite set of universal principles also undermine the possibility of explaining human action on the basis of a nomological system.[7] The uncodifiability thesis is quite plausible in the light of the serious difficulties that we face when we try to apply abstract, universal principles to concrete, particular cases. On the one hand, the vagueness and generality of principles contrast with the uncontrollable diversity of particular situations that defies any set of abstract rules. As a result, it is often possible to arrive at different conclusions departing from the same principle. On the other hand, there are also cases where equally valuable principles may contradict one another. Now, if an agent cannot act on the basis of a precise set of universal principles, the psychologist or sociologist cannot, *a fortiori*, explain her action in the form of deductions from universal principles or laws. As a result, generalizations couched in an intentional vocabulary do not support accurate predictions.

We can now raise the following question: does the explanation of human action involve *ceteris paribus* laws? It is not easy to answer this question, because putative *ceteris paribus* laws have, in the human sciences, an indefinite status; they are a borderline case between, on the one hand, non-scientific rules of thumb or rough generalizations, useful in our everyday practices, and, on the other hand, the laws of the natural sciences.

We could say that the concept of law is a "family-resemblance" concept in the sense that it identifies different kinds of lawlike generalizations without being able to reduce them to a well-defined set of defining characteristics. First, there are exceptionless laws. Some argue, as we have seen, that even the most basic laws of physics are not exceptionless; for instance, magnetic forces may create exceptions to the law of gravitation. However, and since apparent excep-

[7] See also Child 1993.

tions to physical laws may be explained away with the help of other physical laws, we may consider the fundamental laws of physics as exceptionless. Second, there are the non-strict laws of the natural special sciences. Third, there are laws of the human special sciences which are not couched in an intentional vocabulary. They are vulnerable to forms of interference that are typical of the human sciences; for instance, ethical or cultural norms may interfere with well-established economic laws. Fourth, there are generalizations couched in an intentional vocabulary; they are such a pale echo of the basic laws of physics that it becomes disputable to call them laws. Their exceptions are ineliminable because there is not a theory able to deal with the effects of internal interference.

4

Now we must address a further problem: by claiming that the status of scientific generalizations is not the same in the natural and in the human sciences, are we not contributing to a devaluation of the human sciences and returning to an old-fashioned methodological dualism in the realm of science? The answer, as we shall see, is no.

According to the position defended in this article, the explanation of human action is based to a large extent on rationality assumptions. When we want to understand an action, we want typically to understand the reasons that motivated the action, and a mere reference to behavioural regularities is not enough. Human actions are, in most cases, the result of deliberations. It would be quite odd for an agent to take decisions on the basis of her previous behaviour and respective regularities; for the same reason, the explanation of human action cannot abstract totally from the deliberative process that originated a particular action. If our decisions and our self-understanding require the adoption of a deliberator's stance, we cannot explain satisfactorily the human action relying only on observable regularities.[8]

However, one must also recognize that generalizations play an important role in the human sciences. In opposition to strict naturalistic perspectives that downplay the role of rationality assumptions

[8] For an elaboration of this point, see Kim 1998.

in the explanation of human action and hermeneutic approaches that neglect the role of generalizations, we should attempt to articulate rationality assumptions and generalizations in the human sciences.

A prominent supporter of the so-called "simulation theory" in philosophy of mind and cognitive science, Robert Gordon, offers us a good example of such an articulation. The simulation theory is an account of our folk-psychological ability to explain and predict an agent's behaviour on the basis of an intentional vocabulary that includes notions like 'belief', 'desire' or 'intention'. According to a particular view of folk psychology, the so-called "theory-theory", folk psychology is a protoscientific theory, to the extent that it relies on generalizations that, by establishing correlations between thoughts and actions, allow us to predict human behaviour. Simulationists dispute the idea that our folk-psychological ability to explain, interpret and predict human behaviour is based on theoretical knowledge. Several objections may be addressed to the theory-theory. For instance, it is often argued that it is implausible to attribute the relevant theoretical knowledge to small children who are quite good at explaining and predicting actions. The idea that the explanation of human action follows universally shared principles is also controversial. For our purposes, however, the most significant objection against theory theory is based on the vagueness of the putative laws of folk psychology. As Goldman (1995) and Gordon (1995) point out, the laws of folk psychology are condemned to be vague because of the presence of *ceteris paribus* clauses. In fact, the number of interfering factors that can undermine predictions made on the basis of folk-psychological generalizations is so high that the explanatory relevance of such generalizations becomes questionable. As an alternative, simulation theorists claim that we should use ourselves as models for the interpretation of an agent; the proposal is to interpret other minds by projecting our emotions, motivations and deliberating processes.

In this context, Gordon's suggestion that the vagueness of *ceteris paribus* clauses cab be corrected through the method of empathy or simulation is particularly relevant. He rightly points out that by transposing ourselves into other people we can have access to reasons or factors that can interfere with established regularities, enabling us to explain and predict exceptions to reliable generalizations.

> How does one know how to recognize atypical situations or to expand the *ceteris paribus* clause? An answer is ready at hand. As long as one applies these generalizations *in the context of practical simulation*, the unspecifiable constraints on one's own practical reasoning would enable one to delimit the application of these rules. This gives one something to start with: as one learns more about others, of course, one learns how to modify these constraints in applying generalizations to them. (Gordon 1995: 67)

The same point can be made with the help of another distinguished supporter of simulation theory, Jane Heal (2003: 45-62). She rightly points out that the epistemic status of a belief depends on its relations with other, relevant beliefs. As a result, the interpretation of an agent requires sensitivity to the beliefs that an agent considers relevant to a given belief. But how can we identify the relevant beliefs? Certainly not by theory-building, because there is not a theory of relevance. In order to identify relevant beliefs, we have to rely on our own judgements of relevance and to interpret the agent on the basis of such judgements.

This appeal to simulation theory may seem to be incoherent with the Davidsonian premises that I presented above. Simulationists ground our mindreading abilities in psychological mechanisms, and a Davidsonian approach to the interpretation and explanation of human action is based, on the contrary, on rationality assumptions or in the famous Principle of Charity, according to which the interpretation of other agents must necessarily obey to presuppositions of truth and rationality.

However, the method of empathy or simulation is not incompatible with the Principle of Charity. In its initial formulations, the requirement of charity seemed to be a strong normative principle, based on an idealized assumption of truth and rationality in the interpretation of other people's utterances and actions. However, Davidson's later reflections on the Principle of Charity make it clear that the principle combines norms of rationality with considerations of psychological plausibility. The main point is not to maximize truth and rationality, but to avoid the attribution of unexplainable errors. Simulation may involve rationality assumptions to the extent that it reconstitutes the deliberative process of the interpretee; and charity is not a purely normative principle, because it can accommodate psychological intuitions.

Rui Silva
University of the Azores and LanCog Group

References

Cartwright, N. 1997. Where do laws of nature come from? *Dialectica* **51**, 65-78.
Cartwright, N. 2002. In favor of laws that are not *ceteris paribus* after all. *Erkenntnis* 57, 425-439.
Child, W. 1993. Anomalism, uncodifiability, and psychophysical relations. *Philosophical Review* **102**, 215-245
Davidson, D. 1980. *Essays on Actions and Events*. Oxford: Clarendon Press.
Davidson, D. 1984. *Inquiries into Truth and Interpretation*. Oxford: Clarendon Press.
Davidson, D. 1995. Thinking causes. In *Mental Causation*, ed. by J. Heil and A. Mele. Oxford: Clarendon Press, 3-17.
Davidson, D. 2004. *Problems of Rationality*. Oxford: Clarendon Press.
Earman, J., and Roberts, J. 1999. *Ceteris paribus*, there is no problem of provisos. *Synthese* **118**, 439-478.
Earman, J., Roberts, J. and Smith, S. 2002. *Ceteris paribus* lost. *Erkenntnis* **57**, 281-301.
Fodor, J. 1987. *Psychosemantics*. Cambridge, MA: The MIT Press.
Fodor, J. 1991. You can fool some of the people all of the time, everything else being equal; Hedged laws and psychological explanations, *Mind* **100**, 1991, 19-34.
Goldman, A. 1995. Interpretation psychologized. In *Folk Psychology*, ed. by M. Davies and T. Stone. Oxford: Blackwell, 74-99.
Gordon, R. 1995. Folk psychology as simulation. In *Folk Psychology*, ed. by M. Davies and T. Stone. Oxford: Blackwell, 60-73.
Heal, J., 2003. *Mind, Reason, and Imagination: Selected Essays in Philosophy of Mind and Language*. Cambridge: Cambridge University Press.
Hempel, C. 1988. Provisos: A problem concerning the inferential functions of scientific laws. In *The Limits of Deductivism*, ed. by A. Grünbaum and W. Salmon. Berkeley, CA: The University of California Press, 19-36.
Horgan, T. and Tienson, J. 1996. *Connectionism and philosophy of psychology*. Cambridge, MA: The MIT Press.
Kim, J. 1993. *Supervenience and Mind*. Cambridge: Cambridge University Press.
Kim, J. 1995. Can supervenience and 'non-strict laws' save anomalous monism? In *Mental Causation*, ed. by J. Heil and A. Mele. Oxford: Clarendon Press, 19-26.
Kim, J. 1998. Reasons and the first person. In *Human Action, Deliberation and Causation*, ed. by J. Bransen and S. Cuypers. Dordrecht: Kluwer, 67-87.
Kincaid, J. 1996. *Philosophical Foundations of the Social Sciences*. Cambridge: Cambridge University Press.
Kincaid, J. 2004. There are laws in the social sciences. In *Contemporary Debates in Philosophy of Science*, ed. by C. Hitchcock. Oxford: Blackwell, 168-185.
Lange, M. 1993. Natural laws and the problem of provisos. *Erkenntnis* **38**: 233-248.
Lange, M. 2002. Who's affraid of ceteris-paribus laws? Or: How I learned to stop worrying and love them, *Erkenntnis* **57**, 131-147.
McDowell, J. 1998. *Mind, Value and Reality*. Cambridge, MA: Harvard University Press.
Pietroski, P. and Rey, G. 1995. When other things aren't equal: Saving ceteris paribus laws from vacuity. *The British Journal for the Philosophy of Science* **46**,

81-110.
Roberts, J. 2004. There are no laws of the social sciences. In *Contemporary Debates in Philosophy of Science*, ed. by C. Hitchcock. Oxford: Blackwell, 151-167.
Schiffer, S. 1991. *Ceteris paribus laws*. *Mind* **100**, 1-17.
Woodard, J. 2000. Explanation and invariance in the special sciences. *The British Journal for the Philosophy of Science* **51**, 197-254.
Woodard, J. 2002. There is no such thing as a ceteris paribus law. *Erkenntnis* **57**, 27-52.

Metaphysical Analyticity

Célia Teixeira
University of Lisbon and LanCog Group

BIBLID [0873-626X (2012) 34; pp. 869-888]

1 Introduction

The notion of truth in virtue of meaning has played a prominent role in the moderate empiricist account of the a priori. The central tenet of empiricism is the claim that sense experience is the only source of knowledge about the world. However, moderate empiricists also maintain that the way we know certain truths (viz., a priori truths) is different from the way we know empirical truths. In order to conciliate both these claims, empiricists try to show that a priori knowledge, though genuine enough in its own terms, is less substantial, less world-involving than knowledge acquired through experience – the notion of analyticity is used for such a purpose. But the notion of analyticity has been understood in many different ways, some of them unsuitable to be used in an account of the a priori, others more suitable.[1] This is not to say that these notions do not have anything in common. There is (or seems to be) a certain semantic phenomenon that they all endeavour to capture. Whether this phenomenon is real or merely apparent, or whether we manage to capture it with some notion of analyticity is not something I will explore here. My only purpose is to discuss a notion of analyticity that has played a prominent role in the moderate empiricist account of the a priori; namely the notion of truth in virtue of meaning – the so-called *metaphysical notion of analyticity*.[2] Here is how A. J. Ayer characterizes this notion of analyticity:

[1] See Swinburne 1975 for a survey of some of such notions.

[2] The distinction between metaphysical and epistemological notions of analyticity has been introduced by Boghossian (1997).

> I think that we can preserve the logical import of Kant's distinction between analytic and synthetic propositions, while avoiding the confusions which mar his actual account of it, if we say that a proposition is analytic when its validity depends solely on the definitions of the symbols it contains, and synthetic when its validity is determined by the facts of experience. (1946: 33)

If there is such a thing as truth in virtue of meanings alone, and if a priori knowledge is merely knowledge of analytic truths (in this sense), the possibility of a priori knowledge becomes less mysterious: a priori knowledge is merely knowledge of semantic facts (or of conceptual relations between our concepts), and hence purportedly not as substantial or world-involving as empirical knowledge.

However, the metaphysical notion of analyticity has been under a cloud of suspicion ever since Quine (1951)[3] famously attacked its intelligibility. Such an attack has been reinforced by Boghossian (1997) and more recently by Williamson (2007). Nonetheless, the notion is still used in many philosophical circles. My aim is to reinforce this sceptical stance. The paper will proceed as follows. I start by briefly motivating the metaphysical notion of analyticity as well as the main reason to reject it. After this I consider the case of logical truths. I argue that the fact that logical truths are true under all interpretations that preserve their logical form does not in any way provide a reason to think that they are made true by their meanings alone, or that they are somehow less substantial than empirical truths. I then move on to paradigmatic cases of analytic truths, the so-called conceptual truths, to conclude that no reason has been given to regard them as true in virtue of meanings alone, and *a fortiori*, as less substantial or world-involving than empirical truths. I then consider and reject the claim that the analytic/synthetic distinction should instead be regarded as a distinction between two types of truth. I claim that if we are to make sense of the metaphysical notion of analyticity we have to have a less literal understanding of the notion of truth in virtue of meanings alone. I provide such an understanding by drawing from two-dimensionalist approaches to semantics, and I propose a new definition of analyticity according to which a statement expresses an analytic truth iff it remains true no matter how the actual world turns out to be. I argue that though perfectly intelligible, it is not

[3] See also Quine (1935) and (1954).

clear whether this notion captures any distinctive semantic phenomenon deserving of the name 'analyticity', and that even if it does, this in no way shows that analytic truths are somehow less substantial or cognitively demanding than a posteriori truths – as most empiricists have taken them to be.

2 Truth and meaning

In this section I outline the main reason to reject the metaphysical notion of analyticity. But before we proceed let us star by briefly motivating the notion. According to the metaphysical notion of analyticity a statement is analytic if and only if it is true in virtue of its meaning alone. First, let us agree to put semantic scepticism aside. Despite Quine's stance regarding meaning, I will assume, for the sake of the argument, that statements have meanings and that they express contents or propositions. Let us now take the following two statements:

(1) Brutus killed Caesar or Brutus did not kill Caesar.
(2) Snow is white or snow is not white.

Whatever is responsible for the truth of (1) seems to be what is also responsible for the truth of (2). Since the only things that both statements have in common are the logical words 'or' and 'not', it is tempting to claim that the truth of such statements depends entirely on the meaning of those words. Thus, we could claim that such statements are clear cases of analytic statements in the metaphysical sense: they are both true in virtue of their meaning alone.

Now let us take the following statements:

(3) Bachelors are unmarried.
(4) Vixens are female foxes.

These statements are also thought to be true in virtue of their meaning alone, but the reason for this is slightly different from the one above. In this case, these statements are thought to be definitional in nature in the sense of giving the meaning of a certain word. (3) is supposed to give us the meaning of 'bachelor', and (4) the meaning

of 'vixen'. But if such statements only determine meanings, then it seems that they are in virtue of their meaning alone.

Now, the main reason to question the intelligibility of the metaphysical notion of analyticity is the fact that it seems to go against a central truism regarding the relation between truth and meaning. This truism can be stated as follows: for every statement S, S is true if and only if (i) S means that p (or expresses the content that p) and (ii) p is the case. We may call this the *Tarskian Truism*.[4] It is this apparent clash with the Tarskian truism – hereafter (TT) for short – that led Quine to claim that '(...) it is nonsense, and the root of much nonsense, to speak of a linguistic component and a factual component in the truth of any individual statement' (1951: 42). Meanings seem necessary for the truth of a statement. For instance, the statement 'Snow is white' is true, but if instead of expressing the content that snow is white it expressed the content that snow is blue, it would have been false. The problem here is with the claim that meanings *alone* make statements true. To say that the statement 'Snow is white' is true in virtue of saying that snow is white is just part of the explanation. For this statement to be true it also has to be the case that snow is white. The truth of a statement is a function of its meaning and the way the world is. Meanings do not seem to make things what they are, statements only express propositions that hold or not depending on the way things are. However, according to the metaphysical notion of analyticity, the way things are seems irrelevant to the truth of analytic statements. That is, the truth of such statements seems to be exclusively determined by what they express and independently of the things they express being or not the case. But, as Boghossian nicely puts it: 'How could the *mere* fact that S means that p make it the case that S is true? Doesn't it also have to be the case that p?' (Boghossian 1997: 335).

Despite this apparent clash with (TT), many still hold that there is nothing objectionable with this notion of analyticity. Those who maintain it are thus left with two options: either (i) reject (TT) or (ii) interpret the notion of truth in virtue of meaning in a way that does not violate (TT). I will assume that option (i) is not viable: without some independent and principled way of rejecting (TT) such

[4] This is what Boghossian (1997) calls the 'meaning-fact truism'.

an option should be regarded as a non-starter. We are thus left with option (ii) if we want to defend the intelligibility of the metaphysical notion of analyticity. The question now is whether there is a way of making sense of this notion that does not violate (TT). My aim in the following sections is to explore such ways. I will argue that none is forthcoming.

3 Logical truths

Let us consider again statements (1) and (2) above. As we saw, (1) and (2) seem to be true in virtue of the same thing, namely, the meaning of the logical words 'or' and 'not'. But if they are both true in virtue of the meaning of the logical words, does it follow that both statements are about the same thing, namely the meaning of 'or' and 'not'? Remember that according to (TT), a statement S is true iff (i) S means that p (or expresses the content that p) and (ii) p is the case. But if both statements (1) and (2) are true in virtue of the same fact: the meaning of 'or' and 'not', then they should express the same proposition or content (the p must be the same). However, they are about different things – (1) is about Brutus and Caesar and (2) is about snow – and thus cannot express the same proposition (the same p). To claim that they are true in virtue of the same thing – viz. the meaning of the logical words – seems to be in clear violation of (TT); and we have agreed to rule out any attempt to make sense of the metaphysical notion of analyticity that violates (TT).

To make this point a bit more vivid, let us consider the following Portuguese statement:

(2') Ou a neve é branca ou a neve não é branca.

This statement has the same meaning as (2) above, and it would be plainly wrong to translate it as having the same meaning as (1). In this case, statements (2) and (2') are clearly true in virtue of the same thing, for they express the same content (the same p), and p is the case. But since (1) expresses a different proposition or content from the one expressed by (2) and (2'), it cannot be true in virtue of the same p obtaining, on pain of violating (TT).

At this point, a natural move to support the claim that logical

truths are true in virtue of their meaning alone would be to appeal to the standard model-theoretic account of logical truth as truth under all interpretations that preserve logical form. We could then agree that (1) and (2) have different meanings, and maintain that what makes them true is the very same fact: the meaning of the logical words alone.

However, the fact that some truths are true under all interpretations that preserve their logical form does not show that what makes them true is the meaning of the logical words as opposed to the world.[5] For, as we saw, (1) and (2) have different meanings, thus expressing different contents, and so, by (TT), must be true in virtue of those contents obtaining.[6] Moreover, the fact that statements like (1) and (2) both share the following logical form,

(*) P or not P[7]

provides no reason to think that what makes then true is their logical form as opposed to the world. After all, logical forms can also tell us something about the world. For example, (*) could tell us that the world cannot be such that things could happen and not happen at the same time, or that we could have a certain fact and not have it at the same time. In this way we could claim that logical forms, somehow, represent the logical structure of the world.

To make this point vivid, take the following statement:

(5) Schrödinger's cat is dead or is not dead.

[5] See Williamson 2007:64 for a similar claim.

[6] This is not to say that we cannot have different statements with different contents being true in virtue of the same thing. For example, the statements 'Snow is blue or water is H_2O' and 'Snow is green or water is H_2O' express different propositions, but are both true in virtue of the same thing: water being H_2O. But this is no counter-example to my claim, for they are still true in virtue of their contents obtaining.

[7] Notice that we cannot say that what is true in virtue of its meaning alone is the general schema (*) rather than its instances, for (*) is not a statement and thus does not have a meaning.

It is because (5) says something about the world that it poses a problem for our interpretation of quantum mechanics. Nevertheless, (5) is also an instance of (*), and so something that would be true, according to the metaphysical notion of analyticity, in virtue of its meaning alone. And if it were true in virtue of its meaning alone we would not even consider what it says about Schrödinger's cat to determine its truth. Its truth would immediately follow from the meaning of its logical words. This is not, of course, to say that (5) could be false. We still think (or at least most of us) that (5) must be true. But this is not because of the meaning of its logical words, but rather because its logical form mirrors a certain modal structure of the world. Or consider Graham Priest's dialetheism[8] — i.e., the view that there are true contradictions. Priest appeals to several examples to argue that some statements of the form 'Q and not-Q' might be true. The intelligibility of his arguments crucially depends on the fact that the truth or falsity of logical statements is not entirely determined by the meaning of the logical words, but rather depends on the uniform assignment of meaning to each word in a statement in a given context. The fact that some truths are true under all interpretation (or false under all interpretation) that preserves their logical form does not show that what makes them true (or false) is the meaning of the logical words as opposed to the world.

Moreover, as Williamson 2007:64 notes, from the perspective of compositional semantics synthetic truths such as,

(6) Kripke is a philosopher or Kripke is not a man.

are true in the same way as logical truths like (1) and (2) are. Namely, they are true because one of the disjuncts is true. So, if a statement like (1) were truth in virtue of the meaning of its logical words alone, so would (6). But (6) is clearly not true in virtue of the meaning of its logical words alone. Thus, (1) is also not true in virtue of the meaning of its logical words alone. What makes us think that it is true in virtue of its meaning alone is the fact that, contrary to (6), (1) expresses a necessary truth. But the fact that a statement expresses a logically necessary truth in no way entails that it is true in

[8] See, e.g., Priest 1998.

virtue of its meaning alone.[9] Statements that express a logically necessary truth (in the narrow sense) are statements that are true under all interpretations that preserve their logical form, and if this latter property cannot be used to elucidate the notion of truth in virtue of meaning alone, neither can the former.

What makes a statement true or false is what the statement is about. Or to use Aristotle's famous dictum about truth: 'to say of what is that it is not, or of what is not that it is, is false, while to say of what is that it is, or of what is not that it is not, is true.' Statements (1), (2), (5) and (6) are clearly about different things. Thus, they have different truth conditions. Their truth must, therefore, depend on whether such conditions are met. All statements are true, when they are, because their truth conditions are satisfied. And such truth conditions result from a uniform assignment of meaning to each word in the statement in a given context. Logical truths are no different in this respect. The only difference is that logical truths are true under all interpretations that preserve their logical form. But this in no way means that they are true in virtue of the meaning of the logical words alone. Such truths are as much about the extra-linguistic world as any other truth – or at least as any other truth that is not explicitly semantic in nature.

Despite Quine being credit as the first to reject the metaphysical notion of analyticity, Bertrand Russell had already objected to a similar notion. In Russell's time the focus was not so much on meanings but rather on truths of thought, but the intuition was the very same – which is also the intuition present in Hume's distinction between 'relations of ideas' and 'matters of fact'. The way Russell expresses his concern about this divide neatly illustrates the upshot so far:

> The belief in the law of contradiction is a belief about things, not about thoughts. It is not, e.g., the belief that if we *think* a certain tree is a beech, we cannot at the same time *think* that it is not a beech; it is the belief that if a tree *is* a beech, it cannot at the same time *be* not a beech.

[9] The claim that necessary truths were true in virtue of their meaning alone was endorsed by the logical positivists, and rejected by Quine. But the logical positivists did not have independent means to show why this was so. The only reasons given to support such a claim result from their account of the a priori as knowledge of analytic truths, and of their subsequent identification of the concept of a priority with that of necessity.

> (…) and although belief in the law of contradiction is a thought, the law of contradiction itself is not a thought, but a fact concerning things in the world. (Russell 1912: 50)

4 Conceptual truths

What about the paradigmatic examples of analytic truths, the so-called conceptual truths, like 'All bachelors are unmarried', are such statements true in virtue of their meaning alone? The argument here seems to be even more straightforward than the argument regarding logical truths. If the statement 'All bachelors are unmarried' is about bachelors and the property of being unmarried, how can it be true in virtue of its meaning alone?

Consider the following sentences:

(7) All bachelors are unmarried.
(8) All bachelors are healthy.

According to the metaphysical notion of analyticity, (7) is true in virtue of its meaning alone, and (8) is true in virtue of both its meaning and the world (assuming that it is true). In both statements the word 'bachelors' refers to bachelors, and not to the meaning of 'bachelors' (whatever that might be). And it is because the word 'bachelor' refers to bachelors that (7) is about bachelors. But, by (TT), for (8) to be true in virtue of its meaning alone, the statement could not be about bachelors but about the meaning of 'bachelors'. A statement in a context is about whatever its constituent terms refer to in that context. But there is no relevant difference that would make 'bachelors' refer to bachelors in (8) and to the meaning of 'bachelor' (whatever that might be) in (7). But if (7) is about the fact that bachelors are unmarried, such a fact, no matter how trivial it might be, must be what makes it true.

Now contrast (7) with:

(9) 'All bachelors are unmarried' means that all bachelors are unmarried.

If (7) were true in virtue of its meaning alone, the very same fact that made (7) true would made (9) true. But contrary to (7), (9) expresses a semantic truth, and thus must be about a different thing. But if these statements are about different things, by (TT), they cannot be true in virtue of the same thing, namely the meaning of 'All bachelors are unmarried'. Therefore, (7) cannot be true in virtue of its meaning alone.

I submit that if there is no other way of making sense of the metaphysical notion of analyticity such a notion should be rejected as incoherent. In the next section we will look at another way of understanding the metaphysical notion of analyticity.

5 Types of truth

A possible way of understanding the metaphysical notion of analyticity would be to claim that the notion of truth in virtue of meaning alone is not so much about the truth-makers of analytic statements — meanings as opposed to 'the facts' — but about a special type of truth. In this way, claiming that analytic statements are true in virtue of their meaning alone and synthetic statements true in virtue of the facts is to be understood as a claim regarding two different types of truth. So the question that we have to face now is whether there is such a thing as analytic truth and synthetic truth.

Williamson 2007:54-8 has presented a simple and very compelling argument to show why there cannot be two distinct types of truth.

Williamson's argument goes roughly as follows. If there were two distinct senses of truth, how could we determine the truth-value of, for example, a conditional in which the antecedent and the consequent were both synthetic truths? We could say that in such cases the statement that would result from this application would also be synthetic and so true in the synthetic sense. But consider the following statements:[10]

(10) Barbara is a barrister.
(11) Barbara is a lawyer.

[10] I am using Williamson 2007 own examples.

(10) and (11) are clear examples of synthetic truths, if true. Now imagine that Barbara is in fact a barrister, and consider the following conditional:

(12) If Barbara is a barrister, then Barbara is a lawyer.

If there are analytic truths, (12) should be one of them — 'barrister' means a lawyer with special qualifications. However, (12) is composed of two synthetic truths. Does this mean that whenever we have a conditional with a synthetic truth for the antecedent and a synthetic truth for the consequent the result is an analytic truth? This is clearly false, as can be easily illustrated:

(13) If Barbara is a barrister, then Barbara is married.

What this shows is that if there are two senses of 'true' it would be impossible to work out the truth-table for something as simple as the material conditional. Whether a conditional is analytically true or synthetically true is not a function of its parts being analytically true or synthetically true. The best we could do is to claim that the material conditional would be analytically or synthetically true. But this would defeat the whole idea of distinct notions of truth. That is, we would in effect be claiming that there is such a thing as absolute truth.

As Williamson 2007 also shows, we would have the same problem when considering the notion of truth-preservation. Valid arguments preserve truth from premise(s) to conclusion. But if there are two distinct notions of truth, what type of truth is preserved? For example, we could have a valid argument with a synthetically true premise and an analytically true conclusion like the following:

Snow is white.
∴ Snow is white or snow is not white.

It thus seems that we have once more failed to make sense of the metaphysical notion of analyticity. Is there any other way of making sense of this notion? In the next section I will consider one last idea.

6 Analyticity and two-dimensionalism

Let me tentatively suggest a final way of making sense of the metaphysical notion of analyticity.[11] The two-dimensional framework approach to semantics has been understood and used in many different ways, but there are some intuitions that seem common to all of them.[12] One such intuition is the following:

> (…) the truth value of an utterance will depend on facts in two different ways: first, the facts determine what is said; second, the facts determine whether what is said is true. (Stalnaker 2006: 295)

By accepting that the facts also determine whether what a statement says is true, the two-dimensional theorist embrace (TT), as they should. The claim that non-semantic facts also determine what is said is the one that I would like to focus on. Let us see if this could be used to make sense of the metaphysical notion of analyticity.

If we think of a possible world as a complete description of the facts, Stalnaker's quotation above might be understood as referring to the two possible roles a possible world might play: that of (i) determining what is said; and that of (ii) determining whether what is said is true. Now, a world that plays the role of determining what is said is considered to be the *actual world*. A world that is not actual is considered as *counterfactual.* The actual world is normally understood as 'our world' — that is, the way things are. And what determines the truth-value of a statement is the way things are. But there are ways things could have turned out so that what is true might have been false.

Consider the following example of a statement that expresses a contingent truth:

(P) Plato is the author of *The Republic*.

[11] I am grateful to David Papineau for pressing me to think about how the two-dimensional account could be used to make sense of the metaphysical notion of analyticity.

[12] It is not my purpose here to articulate and defend any two-dimensional approach, only to see if we can use some of its ideas to give a coherent account of the metaphysical notion of analyticity. For a discussion of the different interpretations of the two-dimensional semantics apparatus, see Stalnaker 2006 and Chalmers 2006.

By prefixing (P) with the 'Actually' operator we get the following statement:

(PA) Actually Plato is the author of *The Republic*.

The resulting sentence, (PA), now expresses a necessary truth – it is true at every possible world in which Plato is the author of *The Republic* at the actual world. This is the standard way of understanding the 'Actually' operator.

But it seems that there is a sense in which (PA) is not really necessary, that there is something contingent about it. This is the intuition that the two-dimensional theorists want to capture, namely that 'it is a contingent matter which possible world is actual' (Davies 2006: 143). Had a different possible world been actual (PA) could have expressed a falsehood. If another world in which Plato is not a philosopher had been actual, (PA) would have been false.

Now consider the statement,

(W) Water is H_2O.

Following Kripke 1980 and Putnam 1975 we can claim that (W) expresses a necessary truth, because 'water' and 'H_2O' both refer rigidly to the same substance. However, had a different possible world been actual, (W) could have been false. For example, if instead of Earth we had Twin-Earth playing the role of the actual world, then 'Water' would rigidly refer to that watery stuff that plays the water-role on Twin Earth. 'Water' would then rigidly refer to XYZ, and (W) would have been false. So with Twin-Earth playing the role of the actual world, (W) is false.

Now consider the statement,

(WW) Water is water.

Contrary to (W), this statement (WW) is true no matter what world is considered as the actual world. For, even if we had Twin Earth playing the role of the actual world, it would still be true that water is water — for XYZ is XYZ.

Corresponding to these two ways of understanding what is said

are said to be two distinct dimensions: on one dimension we have (A) different worlds considered as actual determining what is said; and on the other we have (C) different worlds considered as counterfactual determining what is said. Dimension (C) corresponds to what David Chalmers calls 'secondary intension', and dimension (A) corresponds to he calls 'primary intension'.[13]

Let us now see if this apparatus could help us make sense of the metaphysical notion of analyticity.

According to the metaphysical notion of analyticity a statement is true if and only if it is true in virtue of its meaning alone. We saw that two-dimensionalism embraces (TT): a statement S is true iff S means that p and p. However, if some statements are true no matter what world is considered actual, then, it might be claimed, they are somehow true in virtue of their meaning alone. After all, if a statement is true regardless of how the world happens to be (regardless of which facts obtain), it seems that it is in some sense true in virtue of its meaning alone (though not literally so).

I submit a slightly modified version of the traditional metaphysical notion of analyticity, one that seems similar in spirit but without its shortcomings. Here is such a notion:

> A statement is analytic (in the metaphysical sense) if and only if it is true at a world *w* regardless of whichever world is labelled as the actual world.

In the useful terminology of Martin Davies and Lloyd Humberstone 1980, analytic statements have, thus, the property of being Fixedly Actually true.

Since a necessary truth, in the standard sense, is one that holds at all possible worlds, we might think that all necessary propositions are analytic in this metaphysical sense of analyticity, and, thus, that (W) is analytic, rendering this definition of analyticity unsuitable. However, this is not so. Statement (W) is not fixedly actually true: it does not hold at the actual world, regardless of whichever world is labelled as the actual world. But (WW) is fixedly actually true, and so analytic – and a priori –, as it should.

We could then claim that the only statements that are analytic in

[13] See, e. g., Chalmers 2006.

this new metaphysical sense of analyticity are those that are knowable a priori.[14] Truths of mathematics and logic are all true in respect to the actual world, whichever world is labelled as actual. The same can be said of the paradigmatic examples of analytic truths, the so-called conceptual truths. So, so far, so good.

Now, if an analytic statement in this metaphysical sense is one that is true no matter how the world turns out to be, it might seem that they do not impose any constraints on the world, and that is why we can know them to be true without having to look at the world – that is, a priori. Synthetic truths, on the other hand, do impose constraints on the world, and so they are somehow more world-demanding than analytic ones. It thus seem that we have finally found a robust enough notion of analyticity to be used in the empiricist account of the a priori. In the next section we will see why this is not so.

6.1 Is this new notion any good?

The first worry I would like to consider regarding this revised notion of metaphysical analyticity is whether it truly captures some special semantic phenomenon deserving of the label 'analyticity'.

The analytic/synthetic distinction is a semantic distinction. The analytic statements, in the metaphysical sense, are the ones that are said to be true for semantic reasons alone, the synthetic statements the ones that are true for non-semantic reasons – they are supposed to be true in virtue of extra-linguistic facts, for they express, in Humean terms, 'matters of fact'. Now, according to this revised notion of analyticity, analytic statements hold (when they do) in virtue of the facts, just as any other statement does. The difference is that they are true at every possible world considered as actual. Given this, one could claim that these statements are such that their meaning somehow guarantees their truth, and hence that they are, to a certain extent, true in virtue of meaning (though not in a literal sense). How-

[14] Peacocke 2004 claims that all propositions that are knowable a priori — those that are 'contentually a priori' in his terminology — are fixedly actually true. Notice, however, that Peacocke strongly rejects the metaphysical notion of analyticity as spurious.

ever, this is not clearly so. The property that a statement has of being true no matter how the actual world turns out to be does not seem to be a purely semantic property, as the property of truth in virtue of meaning clearly is. It seems that if the phenomenon at stake in this new notion of analyticity were semantic, the same would be true of synthetic statements. Analytic statements *are* true no matter how the actual world turns out to be; and synthetic statements are *not* true no matter how the actual world turns out to be. But both statements are true in virtue of their disquotational truth conditions obtaining. The difference is that analytic statements, in this sense, will always have their truth conditions fulfilled and synthetic statements will have them only sometimes fulfilled. But then what is at stake here is a distinct type of necessity, rather than a distinct type of analyticity. And in fact, the 'Fixedly Actually' operator was introduced by Davies & Humberstone 1980 to express a different notion of necessity, what was called 'deep necessity' as opposed to 'superficial necessity', and not a notion of analyticity. The fact that a statement always manages to have their disquotational truth-conditions fulfilled does not show that the phenomenon behind it is merely semantic.

A second worry closely related to this one concerns the role that this purportedly metaphysical notion of analyticity can play in the empiricist account of the a priori. A priori knowledge is knowledge that is, in some suitable sense, independent of the subject's sense experience. If possible, the a priori has the puzzling feature of providing us with knowledge about the world without causally relating us to it. The moderate empiricist way out of this puzzle is to claim that a priori knowledge, thought genuine enough in its own terms, is somehow less substantial or less world-involving than a posteriori knowledge. A priori knowledge is then said to be merely knowledge of analytic truths, and analytic truths understood as semantic in nature, as true in virtue of meaning alone, and so less substantial and cognitively demanding than synthetic truths. The problem, as we saw, is that understood literally, we have to agree with Quine and reject this notion of analyticity. But understood less literally, as statements that are true no matter what world plays the role of the actual world, it is not at all clear that the notion can do the job the empiricists need it to do. For example, mathematical truths are true regardless of how the actual world turns out to be. But this fact does not provide any

reason to regard these truths as somehow less substantial than truths that do not hold fixedly actually, unless we already had independent reasons to think that all truths that hold fixedly actually are not substantial, or that mathematical truths are not substantial. Moreover, if we had independent reasons to think that mathematical truths were less substantial than empirical truths, or that truths that hold fixedly actually were less substantial than those that don't, what would be doing the explanatory work would be such reasons themselves, and not this revised notion of metaphysical analyticity.

The dialectic so far is pretty straightforward. Either this purported notion of analyticity fails to capture a semantic phenomenon deserving of the name 'analyticity' or if it does capture it, it fails to do the explanatory job it was supposed to do. In either case, it should be rejected. This is not to say that the notion of fixedly actually true is irrelevant or philosophical uninteresting, quite the contrary. My only contention is in taking it as capturing a metaphysical notion of analyticity that could be used in an account of the a priori.

It is also worth noting that the identification of the notion of the a priori with this purportedly metaphysical notion of analyticity could lead us to prejudge some important questions. For example, someone could claim that the truth expressed by the statement 'I exist' is not a priori, say, because to know to exist I need to resort to some sort of introspective thought process that seems too analogous to sensory experience for it to be regarded as an a priori source of justification. But, if we were to identify the a priori with this purportedly metaphysical notion of analyticity, instead of addressing the question of whether we can know to exist through some a priori source of justification, we would claim that since this is fixedly actually true it must be a priori. But surely this is not the right way to go about determining the epistemic status of a statement.

For example, Chalmers takes 'I exist' to express an a posteriori truth. However, it is fixedly actually true, and so, it should count as a priori on this approach:

> If S_2 is 'I exist', then any utterance of the same expression with the same meaning will be true, so S_2 has a necessary linguistic and semantic contextual intension [it holds Fixedly Actually]. But (somewhat controversially) S_2 is a posteriori, justifiable only on the basis of experience. [...] All these cases are counterexamples to the Core Thesis [the thesis that identifies what is Fixedly Actually true with the a priori].

> All of them are a posteriori and cognitively significant, and many of them seem to be as cognitively significant as paradigmatic expressions of empirical knowledge. (Chalmers 2006: 70)

My concern here is not so much with whether we should count 'I exist' as expressing an a priori truth — for I think that it does. My concern here is with what two-dimensionalism has to offer as an explanation of the a priori. The aim is to explain the a priori. We do that by claiming that what is a priori is analytic in the sense of holding fixedly actually. We are then faced with what seems to be a counter-example, something that we take to be knowable a posteriori but that is fixedly actually true. We then go on to adjust the theory in order to get a notion of analyticity that is coextensive with the notion of the a priori. Such an adjustment shows that we are using the a priori to test our notion of analyticity. And this is exactly the contrary of what we should be doing if the aim were to explain the a priori with the analytic.

Furthermore, it is not clear that we always know the proposition expressed by a statement that is fixedly actually true. For example, what is it that we know when we know that we exist? The statement 'I exist' is true at every possible world, no matter what world plays the role of the actual world. But what I know when I know that I, myself, exists is not something that holds fixedly actually true. If I am right about this, this means that the identification of the a priori with this purported notion of analyticity is not only problematic but false, as there are things that hold fixedly actually that are not knowable a priori. Of course, one could claim that all a priori truths are fixedly actually true – and that might well be true – but we would also need the converse to hold.[15] That is, we would need all truths that hold fixedly actually to be knowable a priori for this notion to be coextensive with the a priori, as any purported notion of analyticity needs to be.

[15] See, for example, Peacocke 2004, who claims that all a priori truths are necessary in this special sense, though not necessary in the metaphysical sense.

7 Conclusion

I have explored several ways of making sense of the notion of truth in virtue of meaning (i.e., the metaphysical notion of analyticity); I argued that understood literally, the notion should be rejected. I then purposed a less literal way of understanding it. However, though in perfectly good order, it is not clear that this purported new notion of metaphysical analyticity captures a purely semantic phenomenon deserving of the label 'analyticity'. And even if it does, it does not do the explanatory work required by the moderate empiricists. So, even if this new notion does capture a purely semantic phenomenon, this is not the kind of phenomenon that the metaphysical notion of analyticity is supposed to capture. Therefore, in either case, we should follow Quine in his repudiation of the metaphysical notion of analyticity. This notion might well be the last dogma of empiricism.[16]

Célia Teixeira
University of Lisbon and LanCog Group

References

Ayer, A. J. 1946. *Language, Truth and Logic.* London: Gollancz.
Boghossian, Paul. 1997. Analyticity. In *Companion to the Philosophy of Language*, ed. by B. Hale and C. Wright. Oxford: Blackwell.
Davies, M. 2006. Reference, Contingency, and the Two-Dimensional Framework. In *Two-Dimensional Semantics*, ed. by García-Carpintero & Macià. Oxford & New York: Oxford University Press.
Davies, M & Humberstone, L. 1980. Two Notions of Necessity. *Philosophical Studies*, 38: 1-30.
Chalmers, D. J. 2006. The Foundations of Two-Dimensional Semantics. In *Two-Dimensional Semantics*, ed. by García-Carpintero & Macià. Oxford & New York: Oxford University Press.
Jackson, Frank. 1998. *From Metaphysics to Ethics: A Defence of Conceptual Analysis*. Oxford: Claredon Press.
Peacocke, Christopher. 2004. *The Realm of Reason.* Oxford: Oxford University Press.
Priest, Graham. 1998. What's so Bad about Contradictions? *Journal of Philosophy*, 95: 410-26
Quine, W. V. 1951. Two Dogmas of Empiricism. *The Philosophical Review,* 60:

[16] I would like to thank David Papineau for stimulating discussions and David Yates for helpful comments on an earlier version of this paper. This research was supported by a Postdoctoral Fellowship from the Portuguese Foundation for Science and Technology (FCT) (SFRH/BPD/47058/2008 — partially supported by 'Fundo Social Europeu' and MCTES).

20-43.
Quine, W. V. 1935. Truth by Convention. In *The Ways of Paradox and Other Essays.* Cambridge, Massachusetts: Harvard University Press, 1966.
Quine, W. V. 1954. Carnap and Logical Truth. In *The Ways of Paradox and Other Essays.* Cambridge, Massachusetts: Harvard University Press, 1966.
Russell, Bertrand. 1912. *The Problems of Philosophy*. Oxford: Oxford University Press.
Stalnaker, R. 2006. Assertion Revisited: On the Interpretation of Two-Dimensional Modal Semantics. In *Two-Dimensional Semantics*, ed. by García-Carpintero & Macià. Oxford & New York: Oxford University Press.
Swinburne, R. G. 1975. Analyticity, Necessity, and Apriority. In *A Priori Knowledge*, ed. by Paul K. Moser. Oxford: Oxford University Press, 1987.
Williamson, Timothy. 2007. *The Philosophy of Philosophy*. Blackwell/Brown Lectures in Philosophy.

On What There Must Be: Existence in Logic and Some Related Riddles

Paulo A. S. Veloso
Federal University of Rio de Janeiro

Luiz Carlos Pereira
Pontifical Catholic University of Rio de Janeiro

E. Hermann Haeusler
Pontifical Catholic University of Rio de Janeiro

BIBLID [0873-626X (2012) 34; pp. 889-910]

1 Introduction

It is part of an old folklore that logic should not have existential theorems or existential validities. One should not prove in pure logic the existence of anything whatsoever; nothing could be proved by means of logic alone to necessarily exist. Whatever exists might not exist. This standpoint has been expressed by several philosophers from different traditions, such as Hume, Kant, Orenstein and Quine.[1]

We now set the stage by examining some issues. Our main question is: "Do we actually have existential theorems in logic?" Two possible attitudes towards this question are as follows.

[1] See for instance, Hume's Dialogue on Natural Religion [Hum92] (Part IX, 189), the introduction to Kant/Jäsche [Kan92], Orenstein [Ore73] (p. 62) and Quine [Qui54].

- A desideratum: logic should not have any existential theorems.
- A fact: logic does not have any existential theorems.

An intuitive explanation for these attitudes can be roughly as follows. Consider an existential assertion, such as "there exist borogroves", or "something is a borogrove". Such an assertion may perhaps be derived from some hypotheses. But, could it be proved, from no hypothesis whatsoever? In this context, one has absolutely no idea about what 'borogrove' is supposed to mean.[2] So, one can appeal only to logical principles. How would such a proof look like? Here, the problem seems to stem from the fact that we do not know what 'borogrove' is supposed to mean. How can one establish some property of something whose meaning is unknown? Well, actually we can: we do know that "something is a borogrove or a non-borogrove". One might contend that this assertion is not really about borogroves: it expresses a general fact.

The above attitudes towards existential theorems in logic lead to the following slogans concerning (First-order) Logic.

1. No existential conclusion without (perhaps alternative) witnesses.
2. No existential conclusion without universal explanation.

Crudely formulated as above, the preceding attitudes and slogans are clearly misplaced and false: first-order logic does have existential theorems. Our slogans seem to originate from the absence of existential commitments in logic. Well, in a sense, logic does make the following commitments.[3]

[2] Note that a definition of 'borogrove' or other elucidations will not be of much help here. First, they would be counted as hypotheses. Second, if one explains 'borogroves' as, say, "those things that are nimsy", one is bringing 'nimsy' into the picture, and we are back to the same question.

[3] These commitments are incorporated in the introduction rule for the existential quantifier.

(*Cnst*) One usually assumes that constant symbols denote elements of the universe; so the existential sentence $\exists z\, z \doteq c$ is valid.[4]

(*Fnc*) Similarly, function symbols represent total functions on the universe; so the existential sentence $\exists x \exists y\, y \doteq f(x)$ is valid.

Now, consider First-order Predicate Logic with no constants or functions. One may perhaps still detect an existential commitment in the fact that Logic usually considers only nonempty universes. Even if the existence of any particular object is contingent, it is necessary that something exists. The laws of logic are abstract in the sense that they do not presuppose particular objects, but they are not abstract in the sense that they do not presuppose objects. This commitment renders valid the existential sentence $\exists z\, z \doteq z$.[5] Under this modest existential commitment the slogan we propose is: Pure Logic cannot establish existence unless unavoidable.[6]

The structure of this paper is as follows. Section 2 considers the simple case of predicate logic without equality. Section 3 examines the case of predicate logic with equality. Section 4 extends these considerations to the case of universal hypotheses (as axioms). In Section 5, we briefly consider the addition of existential hypotheses. Finally, in Section 6, we summarize our results, commenting on them as well as on some possible extensions.

2 Predicate logic without equality

We now examine the case of predicate logic without equality.

We begin with some simple examples of valid and non-valid sentences.

[4] An assertion like "Pegasus is a winged horse" may be paraphrased as "something identical to Pegasus is a winged horse". Also, "Aristotle is Greek" yields "There are Greeks".

[5] It is this modest commitment that allows one to infer $\exists v\, \varphi$ from $\forall v\, \varphi$, e.g. "Every natural is non-negative" yields "Some natural is non-negative".

[6] The easy road to avoid existential theorems is, of course, to allow empty domains: with empty universes, one would have no existential validity. In in this paper, we take the hard road: even if empty domains are not excluded, predicate logic should not have any existential theorems.

1. Given a unary predicate (symbol) p, consider the matrix $p(z) \rightarrow p(z)$. In this case, the existential sentence $\exists z\,(p(z) \rightarrow p(z))$ is valid, and so is the universal sentence $\forall z\,(p(z) \rightarrow p(z))$ valid.

2. Now, given yet another unary predicate (symbol) q, consider the matrix $p(z) \rightarrow q(z)$. In this case, the existential sentence $\exists z\,(p(z) \rightarrow q(z))$ is non-valid, and so is the universal sentence $\forall z\,(p(z) \rightarrow q(z))$ non-valid.

These two examples illustrate the following situations.

1. For the matrix $p(z) \rightarrow p(z)$, we have:

existential sentence	$\exists z\,(p(z) \rightarrow p(z))$	valid
matrix	$p(z) \rightarrow p(z)$	tautological
universal sentence	$\forall z\,(p(z) \rightarrow p(z))$	valid

2. For the matrix $p(z) \rightarrow q(z)$, we have:

existential sentence	$\exists z\,(p(z) \rightarrow q(z))$	non-valid
matrix	$p(z) \rightarrow q(z)$	non-tautological
universal sentence	$\forall z\,(p(z) \rightarrow q(z))$	non-valid

These situations may suggest a simple, albeit tentative, explanation. Now, let us examine some other examples of valid and non-valid sentences.

Given a unary predicate (symbol) p, consider the matrix $p(y) \rightarrow p(z)$. In this case, notice that the existential sentence $\exists y \exists z\,(p(y) \rightarrow p(z))$ is still valid, but the universal sentence $\forall y \forall z\,(p(y) \rightarrow p(z))$ not valid. There is, however, a way out: the variant versions. We will then have the following situation:

$$\text{Existential variant} \qquad \exists y \exists z \left[\begin{array}{c} (p(y) \rightarrow p(z)) \\ \vee \\ (p(z) \rightarrow p(z)) \end{array} \right] \qquad \text{valid}$$

$$\text{Universal variant} \qquad \forall y \forall z \left[\begin{array}{c} (p(y) \rightarrow p(z)) \\ \vee \\ (p(z) \rightarrow p(z)) \end{array} \right] \qquad \text{valid}$$

These two variants are valid, due to the disjunct $p(z) \rightarrow p(z)$; they have the matrix

$$\left[\begin{array}{c} (p(y) \rightarrow p(z)) \\ \vee \\ (p(z) \rightarrow p(z)) \end{array} \right] \qquad \text{valid}$$

Now, given a binary predicate (symbol) r, consider the matrix $r(y,z)$. In this case, neither the existential sentence $\exists y \exists z\, r(y,z)$ nor the universal sentence $\forall y \forall z r(y,z)$ is valid. The situation with their variant versions is as follows:

$$\text{Existential variant} \qquad \exists y \exists z \left[\begin{array}{c} r(y,z) \\ \vee \\ r(z,z) \end{array} \right] \qquad \text{not valid}$$

$$\text{Universal variant} \qquad \forall y \forall z \left[\begin{array}{c} r(y,z) \\ \vee \\ r(z,z) \end{array} \right] \qquad \text{not valid}$$

Now, the disjuncts $r(y,z)$ and $r(z,z)$ are non-valid, and so is non-valid the matrix

$$\left[\begin{array}{c} r(y,z) \\ \vee \\ r(z,z) \end{array} \right] \qquad (\text{e.g. interpret } r := \emptyset)$$

Let us summarize the ideas seen in our examples.

- Given a matrix, $M(y,z)$, we may have the existential sentence $\exists y \exists z M(y,z)$ valid and the universal sentence $\forall y \forall z M(y,z)$ not valid.
- The examples suggest some simple syntactical transformations as follows.

(SM) The singular matrix $M(z,z)$, obtained from the original matrix $M(y,z)$ by replacing every variable by z.

(VM) The variant matrix $M(y,z) \vee M(z,z)$, obtained by weakening the original matrix $M(y,z)$ by the singular disjunct $M(z,z)$.

One can see that the variant matrix follows from the original one:

$$\underbrace{M(z,z)}_{\text{SM}} \vdash^{(\vee \text{I})} \underbrace{M(y,z) \vee M(z,z)}_{\text{VM}} \qquad (\vee\text{-introduction}).$$

Our examples seem to suggest the following conjectures.

1. Weakening does not affect existential sentence:

$$\frac{\exists y \exists z M(y,z)}{\text{valid}} \quad \Leftrightarrow \quad \frac{\exists y \exists z \, [M(y,z) \vee M(z,z)]}{\text{valid}}$$

2. Propositional reduction for existential validity:

$$\frac{\exists y \exists z M(y,z)}{\text{valid}} \quad \Rightarrow \quad \frac{M(z,z)}{\text{tautology}}$$

3. Tautological variant matrix gives universal variant valid:

$$\frac{M(z,z) \vee M(y,z)}{\text{tautology}} \quad \Rightarrow \quad \frac{\forall y \forall z \, [M(y,z) \vee M(z,z)]}{\text{valid}}$$

4. Universal variant yields existential sentence:

$$\frac{\forall y \forall z \, [M(y,z) \vee M(z,z)]}{\text{valid}} \quad \Rightarrow \quad \frac{\exists y \exists z M(y,z)}{\text{valid}}$$

We now examine these four conjectures.

1. The equivalence between $\exists y \exists z M(y,z)$ and $\exists y \exists z [M(y,z) \vee M(z,z)]$ is easy to see (by the rules for $\exists$ and $\vee$).

2. The idea is considering single-element structures (see below).

3. The assertion follows (by $\forall$-introductions: if $\vdash N$ then $\vdash \forall v N$).

4. The assertion follows from the first assertion (since $\forall \vdash \exists$).

Let us take a closer look at the propositional reduction for existential theorems. There are two ways to establish this reduction, by singularization.[7]

- For classical logic, we consider single-element structures. In such a structure, there are no actual quantifiers: $\exists y \exists z M(y,z)$ holds iff $M(y,z)$ holds for the single element iff $\forall y \forall z M(y,z)$ holds.
- Also, one can show that a predicate-logic proof of $\exists y \exists z M(y,z)$ may be reduced to a propositional proof Π_z of $M(z,z)$ followed by $\exists$-introductions:

$$(\exists) \dfrac{(\exists) \dfrac{\begin{array}{c}\emptyset \\ \Pi_z \\ M(z,z)\end{array}}{\exists z M(y,z)}}{\exists y \exists z M(y,z)}$$

Indeed, we can transform a predicate-logic derivation Π to a propositional derivation Π_z by singularizing it. If we replace every variable by z and erase all quantifiers, we transform both $\forall y N(x,y)$ and $\exists y N(x,y)$ to $N(z,z)$. So, connective rules remain unaffected and likewise for the absurdity rules. The quantifiers rules become repeti-

[7] Such singular translations appear in the litterature, e.g. "réduction de genre un" [Cha79 , p.119-123]. They are often used to establish the relative consistency of predicate logic with respect to propositional logic [End72 , p.59].

tions, except for ∃-elimination, which reduces to derivation gluing. In detail:

- For the quantifier introduction rules, we have:

$$\dfrac{\begin{array}{c}\forall \mathrm{v}\, P(\mathrm{u},\mathrm{v}) \\ \Pi \\ R(\mathrm{x},\mathrm{y})\end{array}}{(\forall \mathrm{I})\ \forall \mathrm{y}\, R(\mathrm{x},\mathrm{y})} \quad \Rightarrow \quad \begin{array}{c} P(\mathrm{z},\mathrm{z}) \\ \Pi_{\mathrm{z}} \\ R(\mathrm{z},\mathrm{z}) \\ \| \\ R(\mathrm{z},\mathrm{z}) \end{array} \qquad \dfrac{\begin{array}{c}\forall \mathrm{v}\, P(\mathrm{u},\mathrm{v}) \\ \Pi \\ R(\mathrm{x},\mathrm{w})\end{array}}{(\exists \mathrm{I})\ \exists \mathrm{y}\, R(\mathrm{x},\mathrm{y})} \quad \Rightarrow \quad \begin{array}{c} P(\mathrm{z},\mathrm{z}) \\ \Pi_{\mathrm{z}} \\ R(\mathrm{z},\mathrm{z}) \\ \| \\ R(\mathrm{z},\mathrm{z}) \end{array}$$

- For the elimination of ∀, we have:

$$\dfrac{\begin{array}{c}\forall \mathrm{v}\, P(\mathrm{u},\mathrm{v}) \\ \Pi \\ \forall \mathrm{y}\, R(\mathrm{x},\mathrm{y})\end{array}}{(\forall \mathrm{E})\ R(\mathrm{x},\mathrm{w})} \quad \Rightarrow \quad \begin{array}{c} P(\mathrm{z},\mathrm{z}) \\ \Pi_{\mathrm{z}} \\ R(\mathrm{z},\mathrm{z}) \\ \| \\ R(\mathrm{z},\mathrm{z}) \end{array}$$

- For the elimination of ∃, we have:

$$\dfrac{\begin{array}{cc}\forall \mathrm{v}\, P(\mathrm{u},\mathrm{v}) & \forall \mathrm{v}\, P(\mathrm{u},\mathrm{v}), [R(\mathrm{x},\mathrm{y})]^1 \\ \Pi & \Sigma \\ \exists \mathrm{y}\, R(\mathrm{x},\mathrm{y}) & Q(\mathrm{w})\end{array}}{(\exists \mathrm{E})\ Q(\mathrm{w})}\,{}_1 \quad \Rightarrow \quad \dfrac{\begin{array}{cc} & P(\mathrm{z},\mathrm{z}) \\ & \Pi_{\mathrm{z}} \\ P(\mathrm{z},\mathrm{z}) & R(\mathrm{z},\mathrm{z})\end{array}}{\begin{array}{c}\Sigma_{\mathrm{z}} \\ Q(\mathrm{z})\end{array}}$$

We then have the following proof transformation:

$$\begin{array}{c} \emptyset \\ \Pi \\ \exists \mathrm{y} \exists \mathrm{z}\, M(\mathrm{y},\mathrm{z}) \end{array} \quad \Rightarrow \quad \begin{array}{c} \emptyset \\ \Pi_{\mathrm{z}} \\ M(\mathrm{z},\mathrm{z}) \end{array}$$

We thus have the following characterizations for the provable existential sentences in Classical, Intuitionistic or Minimal Logics without equality $\doteq$.

Theorem 2.1. *The following assertions are equivalent for a matrix* $M(\mathrm{y},\mathrm{z})$.

(P_1ES) the existential sentence is provable in predicate logic: $\vdash_1 \exists \mathrm{y} \exists \mathrm{z}\, M(\mathrm{y},\mathrm{z})$.

(P_1EV) The existential variant is a theorem: $\vdash_1 \exists y \exists z\,[M(y,z) \vee M(z,z)]$.

(P_0SM) The singular matrix is provable in propositional logic: $\vdash_0 M(z,z)$.

(P_1UV) The universal variant is a theorem: $\vdash_1 \forall y \forall z\,[M(y,z) \vee M(z,z)]$.

Thus, the existential sentences $\exists y \ldots \exists z M(y,\ldots,z)$, without equality, that are provable in predicate logic are exactly those having both:

- a universal explanation

($\vdash_1 \forall y \ldots \forall z\,[M(y,\ldots,z) \vee M(z,\ldots,z)]$),

- witnesses ($\vdash_0 M(z,\ldots,z)$).

3 Predicate logic with equality

We now examine the case of predicate logic with equality.

We begin with some simple examples of valid and non-valid sentences.

1. First, consider the matrix $z \doteq z$. In this case, the existential sentence $\exists z\, z \doteq z$ is valid, and so is the universal sentence $\forall z\, z \doteq z$ valid.

2. Next, consider the matrix $\neg z \doteq z$. Then, the existential sentence $\exists z \neg z \doteq z$ is non-valid, and so is the universal sentence $\forall z \neg z \doteq z$ non-valid.

3. Finally, consider the matrix $y \doteq z$. In this case, the existential sentence $\exists y \exists z\, y \doteq z$ is valid, but the universal sentence $\forall y \forall z\, y \doteq z$ is not valid. Now, the universal variant is the sentence $\forall y \forall z\,[z \doteq z \vee y \doteq z]$, which is valid.

So, the situation is much as before. For a propositional explanation, one must take into account the special nature of the equality sign: $\doteq$ is to interpreted as identity. So, it has some properties embodied in the rules and axioms for equality.

Property	Rule	Axiom
Reflexive	$(\rho)\ \dfrac{}{u \doteq u}$	$\forall u\, u \doteq u$
Symmetric	$(\sigma)\ \dfrac{u \doteq v}{v \doteq u}$	$\forall u \forall v \begin{pmatrix} u \doteq v \\ \downarrow \\ v \doteq u \end{pmatrix}$
Transitive	$(\tau)\ \dfrac{u \doteq v \quad v \doteq w}{u \doteq w}$	$\forall u \forall v \forall w \begin{bmatrix} (u \doteq v \wedge v \doteq w) \\ \downarrow \\ u \doteq w \end{bmatrix}$
Congruence (for unary p)	$(\kappa)\ \dfrac{u \doteq v \quad p(u)}{p(v)}$	$\forall u \forall v \begin{bmatrix} (u \doteq v \wedge p(u)) \\ \downarrow \\ p(v) \end{bmatrix}$

The corresponding z-singular matrices are as follows.

Property	Rule	Axiom
Reflexitivity	$(\rho_z)\ \dfrac{}{z \doteq z}$	$z \doteq z$
Symmetry	$(\sigma_z)\ \dfrac{z \doteq z}{z \doteq z}$	$\begin{pmatrix} z \doteq z \\ \downarrow \\ z \doteq z \end{pmatrix}$
Transitivity	$(\tau_z)\ \dfrac{z \doteq z \quad z \doteq z}{z \doteq z}$	$\begin{bmatrix} (z \doteq z \wedge z \doteq z) \\ \downarrow \\ z \doteq z \end{bmatrix}$
Congruence	$(\kappa_z)\ \dfrac{z \doteq z \quad p(z)}{p(z)}$	$\begin{bmatrix} (z \doteq z \wedge p(z)) \\ \downarrow \\ p(z) \end{bmatrix}$

Thus, except for the case of reflexivity, the singular case is trivial.

We may now provide propositional explanations for the above three examples. The situation is as follows:

1. the matrix $z \doteq z$ is the equality axiom $z \doteq z$;

2. the matrix $\neg z \doteq z$ is not derivable from the equality axiom $z \doteq z$;

3. the matrix $y \doteq z$ has singulary version $z \doteq z$, which is (derivable from) $z \doteq z$.

We thus have the following characterizations for provable existential sentences in Classical, Intuitionistic or Minimal Logics with equality $\doteq$ (cf. Theorem 2.1).

Theorem 3.1. *The following assertions are equivalent for a matrix* $M(y,z)$.

(P_1ES) The existential sentence is provable in predicate logic: $\vdash_1 \exists y \exists z M(y,z)$.

(P_1EV) The existential variant is a theorem: $\vdash_1 \exists y \exists z [M(y,z) \vee M(z,z)]$.

(D_0SM) The singular matrix is derivable in propositional logic from the propositional letter $z \doteq z$*:* $z \doteq z \vdash_0 M(z,z)$.

(P_1UV) The universal variant is a theorem: $\vdash_1 \forall y \forall z [M(y,z) \vee M(z,z)]$.

The argument is much as in Section 2. The only difference lies in the propositional reduction.[8] It now becomes as follows:

$$\begin{array}{c} \emptyset \\ \Pi \\ \exists y \exists z M(y,z) \end{array} \quad \Rightarrow \quad \begin{array}{c} z \doteq z \\ \Pi_z \\ M(z,z) \end{array}$$

So, a predicate-logic proof of $\exists y \exists z M(y,z)$ now has the following form:

$$(\exists I)^2 \dfrac{\begin{array}{c} \overline{z \doteq z} \\ \Pi_z \\ M(z,z) \end{array}}{\exists y \exists z M(y,z)}$$

[8] Alternatively, one can simplify the singular matrix $M(z,z)$ to $M(z,z)_\top$ by replacing each occurrence of $z \doteq z$ by $\top$. Then, the propositional reduction becomes $\vdash_0 M(z,z)_\top$.

4 Universal axioms

We now turn to the case of universal axioms as hypotheses.

We begin with simple examples of propositional reduction. Given the matrices $\boldsymbol{P}(\mathrm{v},\mathrm{w})$ and $\boldsymbol{M}(\mathrm{y},\mathrm{z})$, their singular versions are $\boldsymbol{P}(\mathrm{z},\mathrm{z})$ and $\boldsymbol{M}(\mathrm{z},\mathrm{z})$, respectively.

- Assume that the matrices do not have equality and that we have a propositional derivation Π_0 of $\boldsymbol{M}(\mathrm{z},\mathrm{z})$ from $\boldsymbol{P}(\mathrm{z},\mathrm{z})$. We then have a predicate-logic derivation Π as follows:

$$\cfrac{\begin{array}{c}(\forall\mathsf{E})^2\ \cfrac{\forall\mathrm{v}\forall\mathrm{w}\,\boldsymbol{P}(\mathrm{v},\mathrm{w})}{\boldsymbol{P}(\mathrm{z},\mathrm{z})}\\ \Pi_0\\ \boldsymbol{M}(\mathrm{z},\mathrm{z})\end{array}}{\exists\mathrm{y}\exists\mathrm{z}\,\boldsymbol{M}(\mathrm{y},\mathrm{z})}\,(\exists\mathsf{I})^2$$

Hence, $\forall\mathrm{v}\forall\mathrm{w}\,\boldsymbol{P}(\mathrm{v},\mathrm{w}) \vdash_1 \exists\mathrm{y}\exists\mathrm{z}\,\boldsymbol{M}(\mathrm{y},\mathrm{z})$.

- For the case with equality, assume that we have a propositional derivation Π_0 of $\boldsymbol{M}(\mathrm{z},\mathrm{z})$ from $\mathrm{z} \doteq \mathrm{z}$ and $\boldsymbol{P}(\mathrm{z},\mathrm{z})$. We then have a predicate-logic derivation Π as follows:

$$\cfrac{\begin{array}{c}\cfrac{}{\mathrm{z}\doteq\mathrm{z}}\qquad (\forall\mathsf{E})^2\ \cfrac{\forall\mathrm{v}\forall\mathrm{w}\,\boldsymbol{P}(\mathrm{v},\mathrm{w})}{\boldsymbol{P}(\mathrm{z},\mathrm{z})}\\ \Pi_0\\ \boldsymbol{M}(\mathrm{z},\mathrm{z})\end{array}}{\exists\mathrm{y}\exists\mathrm{z}\,\boldsymbol{M}(\mathrm{y},\mathrm{z})}\,(\exists\mathsf{I})^2$$

Hence, $\forall\mathrm{v}\forall\mathrm{w}\,\boldsymbol{P}(\mathrm{v},\mathrm{w}) \vdash_1 \exists\mathrm{y}\exists\mathrm{z}\,\boldsymbol{M}(\mathrm{y},\mathrm{z})$.

Thus, we can see that:

- the propositional derivability of singular matrices $\mathrm{z} \doteq \mathrm{z}, \boldsymbol{P}(\mathrm{z},\mathrm{z}) \vdash_0 \boldsymbol{M}(\mathrm{z},\mathrm{z})$,
- yields the predicate-logic derivability $\forall\mathrm{v}\forall\mathrm{w}\,\boldsymbol{P}(\mathrm{v},\mathrm{w}) \vdash_1 \exists\mathrm{y}\exists\mathrm{z}\,\boldsymbol{M}(\mathrm{y},\mathrm{z})$.

The converse can be seen by propositional reduction (cf. Section 2). If we replace every variable by z and erase all quantifiers, we transform derivations:

$$\begin{array}{c} \forall x \forall v \forall w\, R(x,v,w) \\ \Pi \\ \exists y \exists z\, M(y,z) \end{array} \quad \Rightarrow \quad \begin{array}{c} z \doteq z \quad R(z,z,z) \\ \Pi_z \\ M(z,z) \end{array}$$

So, a derivation of $\exists y \exists z\, M(y,z)$ from $\forall x \forall v \forall w\, R(x,v,w)$ has the following form:

$$(\exists \mathsf{I})^2 \dfrac{\dfrac{}{z \doteq z} \quad (\forall \mathsf{E})^3 \dfrac{\forall x \forall v \forall w\, R(x,v,w)}{R(z,z,z)}}{\begin{array}{c} \Pi_z \\ M(z,z) \end{array}} \Big/ \exists y \exists z\, M(y,z)$$

Now, let the *singular* z *translation* of a set Δ of universal sentences be the set consisting of their singular z matrices: $\Delta_z := \{R(z,\ldots,z) : \forall x \ldots \forall w R(x,\ldots,w) \in \Delta\}$.[9]

We then have the following characterizations for existential consequences of universal sentences in Classical, Intuitionistic or Minimal Logics with equality $\doteq$ (cf. Theorem 3.1).

Theorem 4.1. *Given a set Δ of universal sentences, the following assertions are equivalent for a matrix $M(y,z)$.*

(D_1ES) The existential sentence is derivable in predicate logic from the set Δ of universal sentences: $\Delta \vdash_1 \exists y \exists z M(y,z)$.

(D_1EV) The existential variant is derivable in predicate logic from the set Δ of universal sentences: $\Delta \vdash_1 \exists y \exists z [M(y,z) \vee M(z,z)]$.

(D_0SM) The singular matrix is derivable in propositional logic from $z \doteq z$ and singular z translation of Δ: $z \doteq z, \Delta_z \vdash_0 M(z,z)$.

(D_1UV) The universal variant is derivable in predicate logic from the set Δ of universal sentences: $\Delta \vdash_1 \forall y \forall z [M(y,z) \vee M(z,z)]$.

[9] The axioms for equality (cf. Section 3) form a set Λ of universal sentences. Note that Λ_z is equivalent to $\{z \doteq z\}$.

5 Existential hypotheses

We now briefly consider the addition of existential axioms as hypotheses.

We begin with a simple examples to introduce the ideas. Given a binary relation r, one may represent the fact that a is related to b via r by an arrow pointing from a to b: $a \to b$. Now, let us call an element cyclic when there is an arrow from it to itself: $c \to c$. Also, call an element bicyclic when there is a path of length 2 from it to itself: $b \to d \to b$, for some d. Let us use C and B for the sets of cyclic and bicyclic elements, respectively.

From the existence of cyclic elements, one can establish the existence of bicyclic elements, i.e. $C \neq \emptyset$ yields $B \neq \emptyset$. One can argue as follows. Since there exists cyclic elements, we have some $c \in C$. Now, as $c \in C$, we have $c \to c$; so we also have the path $c \to c \to c$, whence $c \in B$. This argument actually establishes $C \subseteq B$, which is sufficient.[10]

We claim that establishing such inclusion is the only way to proceed. The intuitive idea is that, for all we know, the assumed cyclic element c may be the only element of the universe. More precisely, consider unary predicates p and q. If $\exists \mathrm{u}\, \mathrm{p}(\mathrm{u}) \models \exists \mathrm{z}\, \mathrm{q}(\mathrm{z})$, the we claim that the inclusion $\forall \mathrm{u}\, (\mathrm{p}(\mathrm{u}) \to \mathrm{q}(\mathrm{u}))$ is valid. Indeed, otherwise, one would have a structure $\mathfrak{M}$ where $\mathrm{p} \subseteq \mathrm{q}$ fails, i.e., for some $a \in M, a \in \mathrm{p}$ but $a \notin \mathrm{q}$. Now, consider the sub-structure $\mathfrak{N}$, with universe $N := \{a\}$. In this structure $\mathfrak{N}$, we have $\mathrm{p} \neq \emptyset$ but $\mathrm{q} = \emptyset$, so $\mathfrak{N} \models \exists \mathrm{u}\, \mathrm{p}(\mathrm{u})$ but $\mathfrak{N} \not\models \exists \mathrm{z}\, \mathrm{q}(\mathrm{z})$.

Now, let us examine some examples of existential consequences of existential hypotheses (without equality, for simplicity).

Given matrices $P(\mathrm{u})$ and $M(\mathrm{z})$, consider the sentences $\exists \mathrm{u} P(\mathrm{u})$ and $\exists \mathrm{z} M(\mathrm{z})$.

(D_1ES) Consider the predicate-logic derivability
$\exists \mathrm{u} P(\mathrm{u}) \vdash_1 \exists \mathrm{z} M(\mathrm{z})$.

Here, the assumption is $P \neq \emptyset$; so there is some element in P, call it u . Much as above, for all we know, u may very well be the only element

[10] This kind of argument appears often in Algebra: from the existence of (right) neuters, one establishes the existence of idempotents, by showing that every (right) neuter is an idempotent.

of the universe. As we wish to have $M \neq \emptyset$, we must have $\mathfrak{u} \in M$. This leads to the propositional reduction, which provides $\mathfrak{u}$ as a witness for $\exists z M(z)$.

(D_0M) We have the propositional-logic derivability of matrices:

$P(\mathfrak{u}) \vdash_0 M(\mathfrak{u})$.

This propositional reduction leads to the universal explanation.

(P_1UC) The universal matrix connection is a theorem:

$\vdash_1 \forall \mathfrak{u}\,(P(\mathfrak{u}) \rightarrow M(\mathfrak{u}))$.

Clearly, this universal connection yields the predicate-logic derivability *(D_1ES)*.

We thus have the equivalence between the assertions *(D_1ES)*, *(D_0M)* and *(P_1UC)*.

For a another example, consider the existential sentences $\exists v \exists w R(v,w)$ and $\exists z M(z)$, as well as a set Δ of universal sentences.

(D_1ES) Consider the predicate-logic derivability:

$$\Delta, \exists v \exists w R\begin{pmatrix} v \\ w \end{pmatrix} \vdash_1 \exists z M(z)$$

Now, the assumption is $R \neq \emptyset$. So, there is some pair of elements, say $\mathfrak{u}$ and $\mathfrak{w}$, in $\mathfrak{u}$. Much as above, for all we know, $\mathfrak{u}$ and $\mathfrak{w}$ might very well be the only elements of the universe.[11] As we wish to have $M \neq \emptyset$, we must have $\mathfrak{u} \in M$ or $\mathfrak{w} \in M$. This leads to the propositional reduction, giving $\mathfrak{u}$ and $\mathfrak{w}$ as (alternative) witnesses for $\exists z M(z)$.

(D_0M) We have the propositional-logic derivability of matrices:

$$\begin{matrix} \Delta_v \\ \cup \\ \Delta_w \end{matrix}, R\begin{pmatrix} v \\ w \end{pmatrix} \vdash_0 \begin{pmatrix} M(v) \\ \vee \\ M(w) \end{pmatrix}$$

This propositional reduction leads to the universal explanation.

(D_1UC) We have the universal consequence connecting the matrices:

[11] Recall that we have no constants or functions.

$$\Delta \vdash_1 \forall v \forall w \left[R\begin{pmatrix} v \\ w \end{pmatrix} \rightarrow \begin{pmatrix} M(v) \\ \vee \\ M(w) \end{pmatrix} \right]$$

Clearly, this universal connection yields the predicate-logic derivability (D_1ES).

For the case with equality, to obtain the propositional reduction it suffices to add the appropriate reflexivity axioms.

Theorem 5.1. *The following assertions are equivalent for existential sentences* $\exists u P(u)$, $\exists v \exists w R(v,w)$ *and* $\exists z M(z)$ *and a set* Δ *of universal sentences.*

(D_1ES) Predicate-logic derivability:

$$\Delta, \begin{matrix} \exists u P(u) \\ , \\ \exists v \exists w R\begin{pmatrix} v \\ w \end{pmatrix} \end{matrix} \vdash_1 \exists z M(z)$$

(D_0M) Propositional-logic derivability of matrices:

$$\begin{matrix} u \doteq u \\ , \\ v \doteq v \\ , \\ w \doteq w \end{matrix} \begin{matrix} \Delta_u \\ \cup \\ , \\ \Delta_v \\ , \\ \cup \\ \Delta_w \end{matrix} \begin{matrix} P(u) \\ , \\ R\begin{pmatrix} v \\ w \end{pmatrix} \end{matrix} \vdash_0 \begin{pmatrix} M(u) \\ \vee \\ M(v) \\ \vee \\ M(w) \end{pmatrix}$$

(D_1UC) Predicate-logic derivability of the universal matrix connection:

$$\Delta \vdash_1 \forall u \forall v \forall w \left[\begin{pmatrix} P(u) \\ \wedge \\ R\begin{pmatrix} v \\ w \end{pmatrix} \end{pmatrix} \rightarrow \begin{pmatrix} M(u) \\ \vee \\ M(v) \\ \vee \\ M(w) \end{pmatrix} \right]$$

The method for constructing the propositional reduction and the universal connection should be clear by now. We consider distinct existentially quantified variables (by resorting to alphabetical variants, if necessary), which gives the witnesses. Then, the existential hypotheses give their matrices. For the existential conclusion, we

replace each one of its variables by the witnesses. For instance, consider the predicate-logic derivability

$$\Delta, \begin{matrix} \exists u\, P(u) \\ , \\ \exists v \exists w\, R\begin{pmatrix} v \\ w \end{pmatrix} \end{matrix} \vdash_1 \exists y \exists z\, M(y,z)$$

Its universal connection is as follows:

$$\Delta \vdash_1 \forall u \forall v \forall w \left[\begin{pmatrix} P(u) \\ \wedge \\ R\begin{pmatrix} v \\ w \end{pmatrix} \end{pmatrix} \rightarrow \begin{pmatrix} M(u,u) \\ \vee \\ M(u,v) \\ \vee \\ M(u,w) \\ \vee \\ \vdots \\ M(w,v) \\ \vee \\ M(w,w) \end{pmatrix} \right]$$

We can extend these ideas to the existential formulas derivable from universal formulas. For the case with equality we have the following characterization.

Theorem 5.2. *The following assertions are equivalent for matrices* $R(u,v)$ *and* $M(x,z)$ *and a set* Δ *of universal sentences.*

(D_1S) Predicate-logic derivability of sentences:

$$\Delta, \exists u \forall v\, R\begin{pmatrix} u \\ v \end{pmatrix} \vdash_1 \forall x \exists z\, M\begin{pmatrix} x \\ z \end{pmatrix}$$

(D_1F) Predicate-logic derivability of formulas:

$$\Delta, \forall v\, R\begin{pmatrix} u \\ v \end{pmatrix} \vdash_1 \exists z\, M\begin{pmatrix} x \\ z \end{pmatrix}$$

(D_0M) Propositional-logic derivability of matrices:

$$\begin{array}{l} \mathrm{u} \doteq \mathrm{u} \\ \quad , \\ \mathrm{x} \doteq \mathrm{x} \end{array} \quad \begin{array}{c} \Delta_{\mathrm{u}} \\ \cup \\ \Delta_{\mathrm{x}} \end{array} \quad \begin{array}{c} R\begin{pmatrix} \mathrm{u} \\ \mathrm{u} \end{pmatrix} \\ , \\ R\begin{pmatrix} \mathrm{u} \\ \mathrm{x} \end{pmatrix} \end{array} \quad \vdash_0 \quad \begin{pmatrix} M\begin{pmatrix} \mathrm{x} \\ \mathrm{u} \end{pmatrix} \\ \vee \\ M\begin{pmatrix} \mathrm{x} \\ \mathrm{x} \end{pmatrix} \end{pmatrix}$$

(D_1UC) Predicate-logic derivability of the universal matrix connection:

$$\Delta \vdash_1 \forall \mathrm{u} \forall \mathrm{x} \left[\begin{pmatrix} R\begin{pmatrix} \mathrm{u} \\ \mathrm{u} \end{pmatrix} \\ \wedge \\ R\begin{pmatrix} \mathrm{u} \\ \mathrm{x} \end{pmatrix} \end{pmatrix} \rightarrow \begin{pmatrix} M\begin{pmatrix} \mathrm{x} \\ \mathrm{u} \end{pmatrix} \\ \vee \\ M\begin{pmatrix} \mathrm{x} \\ \mathrm{x} \end{pmatrix} \end{pmatrix} \right]$$

6 Conclusion

We now review our results, commenting on them and on some possible extensions.

Our aim in this paper has been to provide some substance and evidence for the idea that logic should not have existential theorems, by showing that existence in First-order Logic satisfies two general principles:

1. no existential conclusion without (perhaps alternative) witnesses, and
2. no existential conclusion without universal explanation.

This aim has been attained through several theorems relating existential validities to universal validities and propositional reducts.

First, we review the relationships established here regarding existential derivability (not excluding empty universes). Here Δ is a set of universal sentences.

(UH) Existential sentences derivable from universal sentences can be characterized by derivability of the corresponding universal variants.

$$\begin{array}{ll} \text{Existential sentence} & \Delta \vdash_1 \exists z_1 \ldots \exists z_n M(z_1, \ldots, z_n) \\ & \text{iff} \\ \text{Universal variant} & \Delta \vdash_1 \forall z_1 \ldots \forall z_n \left[\begin{array}{c} M(z_n, \ldots, z_n) \\ \vee \\ M(z_1, \ldots, z_n) \end{array} \right] \end{array}$$

(EH) Existential sentences derivable from existential sentences can be characterized by derivability of the corresponding universal connections.

(EH1) Single existential hypothesis: the universal connection from hypothesis to conclusion provides a witness.[12]

$$\begin{array}{ll} \text{Existential sentence} & \Delta, \exists u A(u) \vdash_1 \exists z G(z) \\ & \text{iff} \\ \text{Universal connection} & \Delta \vdash_1 \forall u [A(u) \rightarrow G(u)] \end{array}$$

(EH1) Two existential hypotheses: the universal connection from hypotheses to conclusion provides alternative witnesses.

We can characterize the first-order derivability

$$\text{Existential sentence} \qquad \Delta, \exists u A(u), \exists w D(w) \vdash_1 \exists z G(z)$$

by means of the connection (for Minimal, Intuitionistic and Classical Logic)

$$\text{Universal connection} \qquad \Delta \vdash_1 \forall u \forall w \left[\left(\begin{array}{c} A(u) \\ \wedge \\ D(w) \end{array} \right) \rightarrow \left(\begin{array}{c} G(u) \\ \vee \\ G(w) \end{array} \right) \right]$$

[12] For instance, from the existence of angels one can derive the existence of gods iff one can prove that every angel is a god.

For the case of Classical Logic, we have a stronger characterization by means of alternative connections.[13]

$$\text{Alternative connections} \qquad \Delta \vdash_1 \left[\begin{array}{c} \forall \mathrm{u}\,(A(\mathrm{u}) \rightarrow G(\mathrm{u})) \\ \vee \\ \forall \mathrm{w}\,(D(\mathrm{w}) \rightarrow G(\mathrm{w})) \end{array}\right]$$

For sentences and formulas, we can summarize these results as follows.

(ES) An existential sentence ES gives rise to a universal sentence US, so that

$$\Delta \vdash_1 ES \text{ iff } \Delta \vdash_1 US$$

Thus, we have our first general principle:

no existential conclusion without (perhaps alternative) witnesses.

(EF) For the case of an existential sentence ES together with universal formulas UF , we have a simple quantifier-free formula $UF^* \rightarrow EF^*$, such that

$$\Delta, UF \vdash_1 EF \text{ iff } \Delta \vdash_1 UF^* \rightarrow EF^*$$

Thus, we have our second general principle:

no existential conclusion without universal explanation.

We now comment on these results and on some possible extensions.

True, one could argue that all these results are limited in an important sense: they are obtained for a pure first order language, i.e., for a first order language without constant and function symbols. As mentioned in the introduction, given constant and function symbols, unless we assume some form of partiality, there is no way to avoid two typical existential validities, namely:

[13] For instance, from the existence of angels and demons one can derive the existence of gods iff one can prove that every angel is a god or every demon is a god.

$$(Cnst)\ \exists z\, z \doteq c \quad \text{and} \quad (Fnc)\ \exists x \exists y\, y \doteq f(x).$$

A possible answer to this kind of criticism can be formulated as follows.

- The language of pure logic is really that of first-order logic without constant and function symbols.
- The above existential validities should be viewed – rather than as theorems of logic – as mere expressions of two general semantical desiderata:

  names must denote and functions must be total.

Although some of the results obtained in this paper can be extended to a first-order language with constant and function symbols, we think it is appropriate to bear in mind here the final version of our slogan: Pure Logic cannot establish existence unless unavoidable.

As future work in this connection, we would like to mention two other analogous slogans we are now examining, namely the following ones.

- There are no necessary existential properties.
- There are no necessary possibilities.

The extension of some of the main results of this paper to properties/relations and modalities could very well provide an interesting starting point towards a systematic treatment for the general issue of existence in logic.

Paulo A. S. Veloso
Federal University of Rio de Janeiro

Luiz Carlos Pereira
Pontifical Catholic University of Rio de Janeiro

E. Hermann Haeusler
Pontifical Catholic University of Rio de Janeiro[14]

[14] Research partly sponsored by the Brazilian agencies CNPq and FAPERJ.

References

Chavineau, Jean. 1979. *La logique moderne* (2e édition). Paris: PUF (Coll. "Que sais-je?" 745).
Ebbinghaus, H.-D., Flum, J. & Thomas, W. 1984. *Mathematical Logic*. Berlin: Springer-Verlag.
Enderton, Herbert B. 1972. *A Mathematical Introduction to Logic*. New York: Academic Press.
Hume, David. *Dialogues on Natural Religion*. London: Penguin Classics, 1990.
Kant, Immamuel. *Lógica*. Rio de Janeiro: Tempo Brasileiro (Biblioteca Tempo Universitário 93), 1992.
Mendelson, Elliot. 1966. *Introduction to Mathematical Logic*. Princeton: D. van Nostrand (repr.).
Orenstein, Alex. 1973. On explicating existence in terms of quantification. In *Logic and Ontology*, ed. by Munitz, Milton K. New York: NYU Press.
Quine, W V O. 1954. Quantification in the empty domain. *Journal of Symbolic Logic* 19: 3 (177–179).
Shoenfield, Joseph R. 1967. *Mathematical Logic*. Reading: Addison-Wesley.
van Dalen, Dirk. 1989. *Logic and Structure* (2nd edition, 3rd printing). Berlin: Springer-Verlag.
Veloso, Paulo A. S., Pereira, Luiz Carlos & Haeusler, Edward H. 2009. *Validades existenciais e enigmas relacionados*. Dois Pontos.