

Centro de Filosofia das Ciências
da Universidade
de Lisboa

FCT
Fundação para a Ciência e a Tecnologia
MINISTÉRIO DA CIÊNCIA, TECNOLOGIA E ENSINO SUPERIOR

CONSELHO DE REITORES DAS UNIVERSIDADES PORTUGUESAS
UNIVERSITAS
CRUP



Lógica e Filosofia da Ciência

Editado por
Olga Pombo e Ángel Nepomuceno



Colecção Documenta
CFCUL

© **Olga Pombo e Ángel Nepomuceno, 2009**

Título: *Lógica e Filosofia da Ciência*
Editores: *Olga Pombo e Ángel Nepomuceno*
Colecção Documenta: *Centro de Filosofia das Ciências da Universidade de Lisboa (CFCUL)*

Faculdade de Ciências da Universidade de Lisboa,
Campo Grande, Edifício C4, 3º Piso, Sala 4.3.24
1749-016 Lisboa, Portugal
http://cfcul.fc.ul.pt
cfcul@fc.ul.pt

Capa, Diagramação e Paginação: *João Araújo*
Impressão: *Guide Artes Gráficas Lda.*
Data: *Janeiro de 2009*
Depósito Legal nº.: *288256/09*
ISBN: *978-989-8247-00-1*

Este livro foi realizado no âmbito do projecto integrado - Logics, Language and Unity of Science - Acções integradas 2008-2009, financiado pela CRUP e estabelecido entre a Universidade de Sevilha e a Universidade de Lisboa, nomeadamente no quadro do Centro de Filosofia das Ciências (POCTI-ISFL-20-678), com o apoio e aprovação da FCT e do POCTI e comparticipado pelo fundo comunitário FEDER.

Textos apresentados nas Jornadas Ibéricas III - Lógica e Filosofia da Ciência, colóquio organizado no âmbito do projecto internacional - Logics, Language and Unity of Science - Acções integradas 2008-2009 entre A Universidade de Sevilha e A Universidade de Lisboa (CFCUL), realizadas na Faculdade de Ciências da Universidade de Lisboa, nos dias 27, 28 e 29 de Fevereiro de 2008.

ÍNDICE

Relaciones de inferencia y el modelo lógico de explicación científica 01
Ángel Nepomuceno Fernández

A most artistic package of a jumble of ideas 19
Fernando Ferreira

Limitations of the Predicate Calculus. The case of Donkey Sentences 21
María J. Frápolli

Sistemas expertos y lingüística computacional: estructuras lógico lingüísticas de lenguas antiguas 33
Cristina Barés Gómez

The Universal set – A (never fought) battle between Philosophy and mathematics 53
Reinhard Kahle

Analogia y semejanza en la representación y identificación de los conceptos teóricos como conceptos complejos 67
Francisco J. Salguero Lamillar

El ideal de la ciencia unificada y sus fundamentos 85
M. Carmen Hernández Martín

Beauty in Mathematics: Aesthetic Terms 93
Ulianov Montaño

Instrumentalismo versus realismo. A crise na física do século XX 121
Rui Nobre Moreira

Consideraciones sobre la referencia de los términos de sustancia **171**
Luis Fernández Moreno

Métodos Formales en Computación Bio-inspirada **185**
Mario J. Pérez-Jiménez

Paradigma, acto e sentido em «Sobre a lógica e a teoria da ciência» de Jean Cavaillès **213**
Nuno Miguel Proença

El mito del método y las estrategias del descubrimiento científico. Inducción, abducción, preducción **231**
Andrés Rivadulla

The beginning of Scientific Controversies. Some Epistemological and Logical Consequences **247**
Hassan Tahiri

«And then there were none» y el razonamento abductivo. El poder esclarecedor de la lógica **267**
Corina Yoris-Villasana

RELACIONES DE INFERENCIA Y EL MODELO LÓGICO DE EXPLICACIÓN CIENTÍFICA*

Ángel Nepomuceno Fernández
Universidad de Sevilla

1. Introducción

Una confusión relativamente frecuente deriva de una falta de clarificación de las nociones de "inferencia" y "deducción", hasta el punto de proclamarse que la lógica se ocupa de estudiar teorías de la argumentación correcta y, al mismo tiempo, aceptar argumentos como correctos sin que los mismos se ajusten a los patrones establecidos por nuestra disciplina. Tal fue el caso, por ejemplo, con los *Elementos* de Euclides, muchas de cuyas "verdades" admitían una demostración irreductible a los patrones silogísticos, paradigmas de la justificación lógica del momento, y por ello se consideraron como conclusiones obtenidas de manera no sistemática (Kneale, 1974). No obstante, la identificación de lógica con silogística es, a su vez, una confusión de este tipo. Asimismo se han producido rechazos a la lógica que se han entendido como rechazos a la propia disciplina, cuando lo repudiado no era más que una teoría lógica particular; como ejemplo baste mencionar los casos de Bacon y Descartes: en último término era la lógica aristotélica (la propia silogística) la que resultaba inservible para lo que debían ser las prácticas científicas, según las concepciones de estos pensadores acerca de la naturaleza del conocimiento científico.

La lógica se suele presentar como una rama de la teoría de la argumentación que se ocupa del estudio de una clase de inferencias; no las más persuasivas, elegantes, etc., sino aquéllas que se consideran correctas en ciertos ámbitos del conocimiento (en la investigación científica, en el razonamiento para la toma de decisiones, en la inteligencia artificial, etc.). Para precisar, aunque sea

* Realizado durante estancia de investigación financiada por el Ministerio de Educación y Ciencia (Gobierno de España), referencia PR2007-0280 en el Centro de Filosofia das Ciências da Universidade de Lisboa, primer trimestre de 2008.

esquemáticamente, diremos que una *inferencia* (o *argumentación*) se presenta como una sucesión de proposiciones, que comienza con un conjunto inicial de las llamadas *premisas* y termina en una denominada *conclusión*; entre ésta y aquéllas se sitúan las que constituyen una *cadena de razones*, aunque en algunos casos esta última es vacía. Los dos extremos (el par <premisas/conclusión>) de una inferencia constituyen su *argumento*; en terminología de Frege, la "función" INFIEREN es diádica y saturable con doble "argumento", a saber, "premisas/conclusión", representable como INFIEREN (premisas, conclusión). Cada inferencia se presenta como una concatenación de argumentos: si no hay cadena de razones, estamos ante la concatenación nula y directamente la inferencia se presenta como un argumento sin más; entonces decimos que la conclusión es una *consecuencia inmediata* de las premisas. En otras ocasiones, la primera proposición de la cadena de razones es consecuencia inmediata de las premisas más, en su caso, otras proposiciones adicionales; de todas ellas, y tal vez de otras necesarias, será consecuencia inmediata la segunda proposición de la cadena de razones, y así sucesivamente.

La corrección de una inferencia está determinada por el modo en que se justifica la inserción de cada proposición en la cadena de razones y el paso de la última de éstas a la conclusión, o bien, directamente por la relación que se presenta entre las premisas y la conclusión, por lo que la corrección de un argumento viene a equivaler a la corrección de la inferencia correspondiente. Precisamente la búsqueda de formas generales de consecuencia inmediata, es decir, de patrones correctos de inferencia, ha sido históricamente una de las tendencias más destacadas en la investigación lógica (Hodges, 1984). En la representación de estos patrones se han venido usando esquemas, lo que culminó con la aparición de los lenguajes formales como herramienta conceptual en los estudios lógicos.

La deducción es una de las formas de inferencia, pero no la única. De hecho hay diversas clasificaciones, de entre las cuales destacamos la propuesta en Peirce (1958), de acuerdo con la cual se consideran tres tipos de inferencia, en función de cómo se relacionan las premisas y la conclusión, a saber

1. Inferencias deductivas. Se trata aquellas inferencias en las que si las premisas fueren verdaderas, entonces la conclusión tiene que ser verdadera.
2. Inferencias inductivas. Son las inferencias en las que si las premisas son verdaderas, entonces es probable que la conclusión también lo sea.
3. Inferencias abductivas. Son inferencias en las que, planteado un problema abductivo, si las premisas (que forman parte de dicho problema) son verdaderas, entonces la hipótesis que se propone (es decir, la conclusión obtenida en este tipo de inferencias) es plausible.

Sin entrar en un estudio detallado de cada uno de estos tipos, hemos de señalar que entre el primero y el tercer tipo, cuando éste último se ha estudiado formalmente, se establece una cierta dualidad. En efecto, el problema abductivo se ha presentado comúnmente en términos de explicación deductiva: el par formado por un conjunto de proposiciones y una proposición constituyen un problema abductivo si se busca completar dicho conjunto para poder deducir tal proposición desde el mismo. Por ello, la conclusión de una inferencia abductiva, junto con el conjunto que constituye una de las premisas (y del problema abductivo en cuestión), será finalmente una premisa adicional (a dicho conjunto), y desde este conjunto ampliado se obtendrá la proposición (segundo elemento del par de premisas inicial) como una conclusión deductiva.

Organizamos el resto del trabajo de la siguiente manera. En el próximo apartado nos ocupamos de la relación de consecuencia lógica en sentido clásico, máximo exponente de la inferencia deductiva; siguiendo los modos de caracterizar la anterior, abordamos el estudio de otras relaciones de consecuencia lógica como variantes deductivas; finalmente estudiamos el modelo lógico de explicación científica, con algunas consideraciones sobre el problema de los requerimientos de deducibilidad. Asimismo, añadimos unas observaciones adicionales y una lista de referencias bibliográficas.

2. Relación de consecuencia lógica clásica

En este contexto nos referimos a lenguajes que pueden formalizar importantes fragmentos de lenguaje ordinario. El lenguaje propio de la lógica de predicados permite, por ejemplo, la formalización del lenguaje especializado de la matemática clásica, de la matemática intuicionista, etc. En cualquier caso, dado un lenguaje formal *L* de predicados de primer orden, la semántica se establece en términos de la teoria de modelos, de manera que *L* se relaciona semánticamente con *estructuras interpretativas del mismo tipo que L*, *L-estructuras* o *L-modelos* (para simplificar, *modelos*). Cada modelo consta de un *universo de discurso* o *domínio*, al cual se impone como única exigencia que no sea vacío, y una *función interpretación* que asigna elementos del dominio a los términos del lenguaje considerados constantes (su papel es análogo al de los nombres propios, denotar entidades del universo de discurso) y predicados y relaciones definidos en el domínio (es decir, subconjuntos del domínio o de la correspondiente potencia cartesiana) a los signos predicativos (monádicos y poliádicos, respectivamente). Tomamos la noción de *satisfacción* como es habitual; para mayor facilidad, solo se asignarán valores de verdad a aquellas fórmulas que pueden formalizar oraciones declarativas del lenguaje ordinário, por tanto a las que no contengan variables libres, también llamadas *sentencias*. Así, la expresión "Antonio estudia filosofia" es formalizable como "F(a)", donde "F" representa el predicado "…estudia filosofia" y "a" es la constante del lenguaje formal representante del sujeto "Antonio"; un *L*-modelo *M* satisface "F(a)" si y solo si (en adelante, "syss") la interpretación de "a" pertenece a la interpretación de "F". La función interpretación, como se observa en este sencillo ejemplo, establece la denotación de los términos constantes (no lógicos) del lenguaje formal *L*, o del lenguaje formal de que se trate. De acuerdo con las reglas de composición de fórmulas, se evalúan las sentencias complejas de la manera que es conocida. Cuando un modelo *M* satisface una sentencia *A* de *L* décimos que "*M* es modelo de *A*" o que "*M* satisface *A*". Si *M* satisface todas y cada una de las sentencias que integran el conjunto *S*, diremos, para abreviar, que "*M* es modelo de *S*" o que "*M* satisface *S*".

La relación de consecuencia lógica clásica es una relación entre conjuntos de sentencias y sentencias del lenguaje formal correspondiente. Así, dados un conjunto de sentencias *S* y una

sentencia *A* de *L*, diremos que *A* es *consecuencia lógica* de *S* (o *S entraña* o *implica lógicamente A*) en sentido clásico syss todos los modelos de *S* son modelos de *A*. Para simplificar, omitiremos "en sentido clásico", salvo que fuere necesario evitar ambigüedad. En lugar de una relación de consecuencia se puede establecer una operación sobre conjuntos de sentencias; en tal caso se define el *conjunto de las consecuencias* de *S* como el de las sentencias de *L* para las cuales se cumple que todo modelo de *S* es modelo de éstas. A lo largo del trabajo usaremos más la primera noción que la segunda, pero se trata de nociones intercambiables ya que, de acuerdo con las definiciones, *A* es consecuencia lógica de *S* syss *A* pertenece al conjunto de las consecuencias de *S*. Un par (*S, A*), un conjunto de sentencias y una sentencia, respectivamente, representa un argumento expresado en *L,* y puede ser el correlato formal de un argumento del lenguaje ordinario, por ello los elementos del par son llamados premisas y conclusión, respectivamente. Un argumento (*S*, *A*) de *L* es *válido* syss *A* es consecuencia lógica de *S*.

La relación de consecuencia lógica tiene algunas propiedades características, o *reglas estructurales*, al margen del lenguaje formal específico que se esté usando, por tanto con independencia de qué signos lógicos posea *L*. Estas son

- Reflexividad. Para todo conjunto de sentencias *S*, cada sentencia elemento de *S* es consecuencia lógica de *S*.
- Monotonía. Si la sentencia *A* es consecuencia lógica del conjunto de sentencias *S*, entonces *A* es consecuencia lógica en sentido clásico de todo conjunto de sentencias que contenga a *S* (es decir, de todo superconjunto de *S*).
- Transitividad. Si una sentencia *A* es consecuencia lógica de un conjunto de sentencias *S* y la sentencia *B* es a su vez consecuencia lógica de *A* (en este caso, del conjuno unitario cuya única sentencia es *A*), entonces *B* es consecuencia lógica de *S*.

Además de las tres reglas estructurales, la relación de consecuencia lógica verifica *compacidad*, de acuerdo con la cual, si se da que la sentencia *A* es consecuencia lógica del conjunto de sentencias *S*, entonces, existe algún sobconjunto finito *S'* de *S* tal que

A también es consecuencia lógica de *S'*. Asimismo cumple la *sustitución uniforme*, pues definida ésta como operación de intercambio de términos no lógicos en fórmulas, con ciertas restricciones, la relación de consecuencia lógica se mantiene tras una sustitución uniforme, tanto en las premisas como en la conclusión.

3. Otras relaciones de consecuencia

Cualquier relación que se defina entre conjuntos de sentencias y sentencias, si verifica las tres reglas estructurales antes relacionadas, es una relación de consecuencia *de clausura*. La relación clásica estudiada es, por tanto, una relación de clausura. En determinados ámbitos tal vez convenga restringir qué conclusiones se deberían obtener de según qué premisas. Así, la exigencia de unas condiciones de relevancia, da lugar a una nueva definición de relación de consecuencia. Para distinguir de la ya estudiada, diremos que la sentencia *A* es *consecuencia relevante* del conjunto de sentencias *S* (*o S entraña relevantemente A*, o *entraña* o *implica relevantemente A*) syss se verifican conjuntamente las dos condiciones siguientes

1. *A* es consecuencia lógica de *S,*
2. *S* es consistente y la negación de *A* también (por tanto, no se puede deducir de *S* una contradicción, ni puede ser conclusión una sentencia universalmente válida).

Naturalmente, se trata de una relación de clausura, que también verifica compacidad, aunque no cumple sustitución uniforme. Makinson (2005) ha estudiado diversas vías para definir nuevas relaciones a partir de las de clausura, usadas como lógicas-puente para tratar de caracterizar otras lógicas, en particular aquellas en cuya relación de consecuencia no se cumple la regla de monotonía. Aquí nos interesamos por tales lógicas-puente, las cuales se establecen a partir de algunas restricciones en la relación de consecuencia y se clasifican en tres grupos

1. Relación de *consecuencia módulo un conjunto de sentencias T*. La sentencia *A* es consecuencia módulo *T* del conjunto de sentencias *S* syss todos los modelos de *S* que sean modelos de *T*,

son modelos de *A*. Ello equivale a que *A* sea consecuencia lógica de *T* y *S* conjuntamente. Así pues, es una relación de clausura que también verifica compacidad, aunque no sustitución uniforme.

2. Relación de *consecuencia módulo un conjunto de reglas R*. En este caso, de manera análoga a las condiciones de relevancia, las exigencias adicionales vienen dadas por un conjunto de reglas *R*. Para un conjunto de sentencias *S* se define *R(S)* como el conjunto de sentencias que son conclusión de alguna premisa (de reglas de *R*) que está en *S* y se dice que *S* es cerrado bajo *R* syss contiene (o es igual a) *R(S)*. Así *A* es consecuencia lógica módulo *R* de *S* syss *A* pertenece a cada superconjunto de *S* que sea cerrado bajo *R*. En este caso, aún tratándose de una relación de clausura, según sea *R* (para ciertas clases de reglas, no estaríamos realmente ante una relación de clausura), no se cumple compacidad, tampoco sustitución uniforme.

3. Relación de *consecuencia módulo un conjunto de modelos M*. Ahora las exigencias adicionales no son más que una restricción de los modelos a considerar. Decimos que *A* es consecuencia lógica módulo *M* de *S* syss cada *L*-modelo del conjunto *M* que satisface *S* también satisface *A*. En este caso, según sea la clase de modelos *M,* tal vez no se cumpla compacidad (un ejemplo: cuando se trata de la clase de los modelos cuyos universo de discurso tienen un número finito de elementos). Tampoco se cumple sustitución uniforme.

Las relaciones módulo y la restringida con condiciones de relevancia son aplicables en la representación de un contexto inferencial, entendiendo por tal todas aquellas circunstancias que permiten determinar lo que realmente se infiere desde un conjunto de premisas dado. En Nepomuceno-Soler (2007) se estudian representaciones de contextos inferenciales con los tres tipos de inferencia señalados. Una cuestión que se plantea es si son definibles cálculos que den cuenta plenamente de las relaciones en cuestión, lo que lleva a otras cuestiones no menos interesantes; más en concreto, si son definibles cálculos que sean correctos y completos (con una semántica como la bosquejada más arriba y respecto de las relaciones de consecuencia de que se trate), qué problemas se plantean acerca de la decidibilidad, complejidad, etc. Los cálculos, en último término,

establecen una relación de *teoremicidad*, entre conjuntos de sentencias y sentencias, es decir, una relación de carácter sintáctico, en lugar de semántico, similar a las de consecuencia. Así, una sentencia *A* es un *teorema en un cálculo lógico* desde las sentencias *S*, los supuestos, syss desde *S* se obtiene *A* mediante sucesivas aplicaciones de los postulados del cálculo.

Aunque es sabido, recordamos que un cálculo lógico es correcto cuando los teoremas obtenidos desde supuestos dados son consecuencia (en el sentido de la correspondiente relación) de tales supuestos y, recíprocamente, es completo cuando cada conclusión de premisas dadas (en el sentido de la relación de consecuencia en cuestión) es un teorema de éstas tomadas como supuestos. A este respecto, para conseguir un cálculo correcto y completo que dé cuenta de una relación de clausura, una condición necesaria es que ésta cumpla compacidad. Ello es así porque, como se muestra en Amor (2006), si un cálculo es completo, entonces se verifica compacidad y si la relación de consecuencia en cuestión no la verifica, la de teoremicidad definida para captarla, tampoco, de donde, por contraposición, es de negar que el cálculo fuese completo.

De acuerdo con lo expuesto, podemos concluir que son diversas las relaciones de consecuencia de clausura y que, en determinados casos, se pierde alguna de las características de la relación en sentido clásico como son la compacidad y la sustitución uniforme. Los casos en cuestión se corresponden con los contextos inferenciales en los cuales se asumen algunos asertos (consecuencia módulo un conjunto de sentencias), reglas adicionales (consecuencia módulo un conjunto de reglas), una clase específica de interpretaciones (consecuencia módulo un conjunto de modelos) o se imponen ciertas condiciones de relevancia. Tratándose, en todos los casos, de relaciones de clausura, podemos convenir en que en tales contextos inferenciales se mantiene un "estilo clásico" de inferencia y, aunque no siempre completos, en todos los casos son definibles cálculos lógicos correctos. En Velasco (2006) se estudia un concepto de *modelo inferencial clásico* como resultado del desarrollo histórico de la lógica que se sintetiza en los siguientes puntos

1. El establecimiento de la lógica proposicional megárico estóica
2. El desarrollo de una dialéctica formal por parte de los lógicos medievales

3. Aparición del cálculo de probabilidades en el siglo XVII
4. Establecimiento definitivo de la moderna lógica matemática a partir de la aparición de la *Conceptografía* de Frege

Asimismo señala que se produce una ruptura entre la lógica matemática y los desarrollos posteriores, desde el comienzo de los años setenta del siglo XX, en los que surge un gran interés por la psicología cognitiva, la inteligencia artificial, la lógica informal, el pensamiento crítico, etc. No obstante, la noción de modelo inferencial es definible en términos del análisis aquí propuesto: cada contexto inferencial que esté regido por una relación de consecuencia de clausura se ajusta al modelo inferencial clásico. Los cálculos lógicos se establecerán atendiendo a los signos lógicos que posea el lenguaje formal de que se trate; la teoremicidad que se dé entre un conjunto de sentencias y una sentencia implicará que ésta es consecuencia (de clausura) de aquéllas, es decir, la exigencia de corrección será irrenunciable, pero la circunstancia de que no sea definible un cálculo lógico completo no implica un cambio de modelo inferencial. Un caso de este tipo se presenta en la lógica de predicados de segundo orden, cuya cuantificación no se restringe a variables de individuo; en este ámbito son definibles cálculos correctos y no completos y sigue siendo considerada (al menos) como una extensión de lógica clásica.

Naturalmente hay otras relaciones de consecuencia que no son de clausura. Se suele presentar como ejemplo de inferencia deductiva que no encaja en el modelo clásico el de las reglas que admiten excepciones. Lo común en estos casos es que se esté argumentando con información incompleta: se obtienen conclusiones a partir de premisas, pero un incremento de información (un aumento de las premisas, en definitiva) podría llevar a un rechazo de la conclusión aceptada inicialmente. Como ejemplo, consideremos la regla "para todo *x*, normalmente, si *x* es un *X,* entonces *x* tiene la propliedad *P*"; dado que *a* sea un *X*, se infiere, de acuerdo con la regla, que *a* tiene la propiedad *P*, pero ello será así en "circunstancias normales", de manera que ante una nueva premisa, que informe acerca de las circunstancias, si éstas dejan de ser normales, se puede llegar al rechazo de la conclusión que se había establecido en principio.

Para este tipo de inferencias las relaciones de consecuencia de clausura no constituyen un correlato formal adecuado, dado que la monotonía característica de esta clase de relaciones constituye una

garantía de que ninguna premisa nueva lleva a un cambio en la conclusión. No obstante, cabe definir relaciones de consecuencia distintas de las de clausura. Sean un conjunto de sentencias *S* y una sentencia *A*, de manera que se verifica

> **1.** *S* y *A* tienen determinada estructura sintáctica, caracterizada por un conjunto de reglas claramente definidas y de carácter decidible.
> **2.** *S* conjuntamente con *A* son consistentes, lo que se puede entender semánticamente como que existe algún modelo que satisface *S* junto con *A*.

En este caso se puede considerar una relación de *consecuencia débil*: *A* sería consecuencia débil de *S* syss se cumplen 1 y 2. Esta relación, establecida adecuadamente la condición 1, no es de clausura. En efecto, aunque esta condición se defina de manera que se verifique reflexividad y transitividad, no se verifica monotonía, pues si *A* es consecuencia débil de *S*, si *S'* se obtiene aumentando el conjunto de sentencias *S* con la negación de la sentencia *A*, entonces *A* no es consecuencia débil de *S'* (por no cumplir 2), por lo que la ampliación de las premisas "no mantiene" la relación de consecuencia. Los sistemas tipo Reiter son sistemas lógicos que se definen mediante un conjunto de reglas específicas para la construcción de una teoría. A continuación bosquejamos cómo se presenta un sistema normal tipo Reiter. Desde un conjunto inicial de sentencias consistente *T,* dada una relación sintáctica *R* entre conjuntos de sentencias (que es transitiva y, además, si *C* es una sentencia consecuencia lógica clásica del conjunto de sentencias *P*, entonces *P* y el conjunto que sólo tiene a *C* como elemento, están en la relación), decimos que una sentencia *A* es *R*-consecuencia del conjunto de sentencias *S* (consistente con *T*) syss se verifica alguna de estas dos condiciones

> **1.** *A* es consecuencia lógica de *S* (en sentido clásico)
> **2.** *S* es un subconjunto no vacío de *T*, el conjunto que tiene *A* como única sentencia está en la relación *R* con *S* y *A* es consistente con *T*

Veamos brevemente el análisis estructural de esta relación

- Reflexividad. Se verifica por la reflexividad misma de la noción clásica, ya que para cada sentencia del conjunto de *S* se verifica la condición 1
- Transitividad. Sean *A* y *B* tales que *A* es *R*-consecuencia de *S* y *B* es *R*-consecuencia de *A*. Entonces se pueden presentar los siguientes casos: (a) *A* es consecuencia lógica de *S* y *B* lo es de *A*; entonces, por ser ésta de clausura verifica transitividad. (b) Se da la relación porque en los dos casos se cumple específicamente la condición 2 de su definición. Entonces, por la transitividad de *R*, *S* y el conjunto que tiene a *B* como único elemento están en la relación *R*; como *B* es consistente con *T* y se dan las demás condiciones establecidas en la cláusula 2, *B* es *R*-consecuencia de *S*. (c) Los casos en que la primera *R*-consecuencia se dé por la cláusula 1 y la segundo por la 2, o (d) recíprocamente, son análogos al anterior.
- Monotonía. Dado que *A* sea *R*-consecuencia de *S* por cumplir la cláusula 2, si *S'* es un superconjunto de *S*, pero *S'* no está en la relación *R* con el conjunto que tiene *A* como único elemento, entonces *A* no es *R*-consecuencia de *S'* . Por tanto, no se cumple la regla de monotonía.

Generalmente, se suelen definir estos sistemas de modo que cumplan una denominada *monotonía cautelosa*. En el ejemplo anterior, basta que la relación sintáctica *R* tomada como punto de partida verifique lo siguiente: si los conjuntos *S* y *P* están en la relación y, asimismo, los conjuntos *S* y *Q;* entonces también lo están los conjuntos "unión de *S* y *P*" y *Q*. Entonces se cumple dicha forma de monotonía, es decir

- Si *A* es *R*-consecuencia de *S* y *B* es *R*-consecuencia de *S*, entonces *B* es *R*-consecuencia de *S* junto con *A*.

Así pues, se pone de manifiesto la existencia de relaciones de consecuencia no monótonas, las cuales podrían formalizar ciertos modos de inferencia que no se ajustan al modelo clásico.

4. Modelo lógico de explicación científica

Nos vamos a referir al denominado *modelo nomológico-deductivo* de explicación, presentado inicialmente en Hempel-Openheim (1948). Básicamente, en este modelo lo que se plantea es que dada(s) una(s) ley(es) científica *L*, junto con las condiciones antecedentes *C*, se explica deductivamente el hecho *E*. Al hecho se le denomina *explanans*, y el *exlanandum* está constituido por *L* y C, mientras que la relación que mantienen *explanans* y *explanandum* debe ser la de una deducción lógica. Es decir, a partir de *L*, para explicar el hecho *E*, basta con hallar *C* de manera que *E* se deduzca desde *L* y *C* tomados conjuntamente. Desde un primer momento, lo que se entiende por deducción es la deducción de la lógica clásica, la cual, según hemos visto más arriba, no es más que la teoremicidad de un cálculo clásico, es decir, definido de manera que capte la relación de consecuencia lógica clásica, como hemos mencionado, una relación de clausura que comúnmente verifica compacidad. Por ello, en términos de las relaciones estudiadas, el modelo nomológico deductivo nos viene a mostrar que la explicación se alcanza cuando contamos con un *explanans* que es consecuencia lógica del *explanandum*.

Los estudios lógicos de la propuesta inicial matizaron la caracterización de los componentes del modelo. Eberle, Kaplan & Montague (1961) destacan que los requerimientos que se exigieron al presentar el modelo no son los más adecuados; estos habrían llegado a permitir que, por ejemplo, asumiendo que no hay sirenas, la ley general "las sirenas son buenas conductoras del calor" y la condición "si la Torre Eiffel (TE) no es una sirena, entonces la TE es una buena conductora de calor" explican el hecho "la TE es una buena conductora del calor". En efecto, en términos de un cálculo deductivo-natural clásico, de dicha ley se obtiene mediante simple especificación de generalización (es decir, eliminación de generalizador universal) que "si la TE es una sirena, entonces TE es una buena conductora de calor", asumiendo el principio de tercero excluido como una ley lógica (demostrable en estos cálculos, de tipo Gentzen), tendremos que "la TE es una sirena o la TE no es una sirena" y, por la ley de casos, se concluye que "la TE es una buena conductora del calor". En este y en posteriores trabajos se refinaron los requerimientos exigibles para evitar que el modelo justificara unas explicaciones tan poco

convincentes. Finalmente se le llamó *covering law model* y a partir de éste hallamos interesantes propuestas como la de Halonen-Hintikka (2005) quienes subrayan la importancia de considerar la totalidad del proceso de explicación, el cual tiene su punto de partida en un conjunto de premisas teóricas, es decir, una teoría base inicial y mediante una lógica de "cuestiones por qué" (*why questions*), responder a las cuestiones consistiría en combinar las premisas iniciales (la teoría base inicial) con aportaciones adecuadas relativas a la situación en la que se plantea la explicación, de una manera tal que el hecho a explicar se haga como inevitable. La condición de deducibilidad se propone ahora más compleja: T[P] representa la teoría base, I[b] las condiciones iniciales y mediante P(b) nos referimos al *explanandum*; entonces se deberían satisfacer las siguientes condiciones:

1. b no ocurre en T[P]
2. P no ocurre en I[b]
3. I[b] no es inconsistente
4. P(b) no se deduce sólo de T[P]
5. P(b) no se deduce sólo de I[b]
6. I[b] debe ser compatible con la teoría T[P]

En opinión de estos autores la mayoría de los contraejemplos aducidos contra el modelo no cumplen con todas estas condiciones. Además del uso del "modelo interrogativo", hacen algunas aplicaciones del "teorema de interpolación" para obtener:

- H[b] se deduce de I[b], donde H[b] es una fórmula que se obtiene al cumplirse los requisitos mediante el teorema de interpolación,
- De T[P] (la teoría base inicial) se deduce "para todo x; si H[x], entonces P(x)" (la *covering law*, ahora separada de la teoría base inicial)
- P no ocurre en H[x]

En la misma línea Schurz (2005) propone también la lógica epistémica de cuestiones como el instrumental lógico adecuado para abordar el estudio de la explicación científica, aunque discrepa en

cuanto al análisis concreto de las "cuestiones por qué" en el marco de Hintikka.

De todo ello se concluye que sería un error tomar la deducibilidad de una manera simplista como único requisito exigible al modelo; se deben considerar otras estipulaciones adicionales para que una teoría y unas condiciones antecedentes constituyan un *explanans* de un *explanandum*. En cualquier caso, el problema de la explicación científica se puede abordar también en términos de inferencia abductiva. En Aliseda (1998, 2006) se investiga cómo el modelo nomológico-deductivo viene a reflejar un proceso de inferencias de carácter abductivo: a partir de una teoría y un hecho a explicar en términos de la teoría, se obtienen las condiciones antecedentes. Entonces estamos ante un proceso en el que las premisas están constituidas por la teoría y el hecho (la fórmula que lo representa) y la conclusión que se obtiene son las condiciones antecedentes. Desde un punto de vista lógico hay varias formas de abducción. En concreto, la abducción explicativa consistente tiene las siguientes características (Aliseda, 2006), para una teoría T y un hecho a explicar E; es decir dado un problema abductivo (T, E): C es una solución explicativa y consistente de (T, E) syss

1. E es consecuencia lógica de T junto con C
2. T junto con C es un conjunto de sentencias consistente
3. E no es consecuencia lógica de T

Fácilmente se constata que estas condiciones constituyen en buena parte un resumen de las indicadas en las propuestas de Halonen-Hintikka (2005). Por otro lado, se han aplicado cálculos deductivos para la búsqueda de soluciones a problemas abductivos; en Mayer, Pirri (1993), Aliseda (1997; 2006) y Nepomuceno (2002) se han aplicado tablas semánticas y en Soler (2005) se aplica un cálculo de resolución dual. En todos estos casos se consideran determinadas características de los cálculos clásicos. Representando la negación de una sentencia A como $\neg A$, el método de tablas semánticas (abreviadamente, la tabla semántica de A, B, por ejemplo, lo expresamos como $TS(A,B)$, es decir, anotando TS y escribiendo entre paréntesis las sentencias o conjuntos de sentencias correspondientes), se aplica de acuerdo con la siguiente propiedad: E es consecuencia lógica de T y C syss la $TS(T,C, \neg E)$ es cerrada (lo que equivale a que

contiene un par de literales complementarios, o, lo que es lo mismo, que contiene una sentencia atómica y su negación). Para que se cumplan los requisitos de la abducción explicativa consistente, se ha de verificar que *TS(T, C)* sea abierta (una tabla es abierta syss no es cerrada) y *TS(T, ¬E)* debe ser abierta; entonces, a partir de esta última tabla, se obtiene el conjunto de sentencias que al añadirlas a la tabla la cerrarían. Este es el conjunto de las soluciones abductivas, que debe ser depurado por procedimientos ajenos al propio cálculo de tablas, para llegar finalmente a las condiciones antecedentes. En el caso de la resolución dual, en lugar de tomar como base las formas normales disyuntivas se toman sus duales, formas normales conjuntivas, y el principio de resolución dual es sintácticamente igual que el estándar, si bien con ese carácter dual desde un punto de vista semántico.

El planteamiento del modelo nomológico-deductivo sirvió como punto de partida de una fructífera investigación en filosofía de la ciencia. La primera aspiración de hallar un modo unificado de abordar estos temas en ciencias empíricas y en ciencias humanas y sociales, antes del renovado interés suscitado por las propuestas de Halonen-Hintikka (2005), fue cediendo paso a una mayor preocupación por aspectos más allá de las cuestiones lógicas implicadas. Una muestra de la variedad de enfoques la hallamos en W.J. González (2002), que recopila una serie de trabajos en torno a las explicaciones causales de Salmon. Tanto en los estudios de condiciones exigibles a este modelo como en los rechazos al mismo, la condición de deducibilidad se ha seguido entendiendo en el sentido de deducibilidad clásica, es decir, en el de una relación de consecuencia lógica clásica.

Concedamos que la búsqueda de explicación lógica se puede considerar como un problema abductivo, según lo indicado anteriormente. Se trata de un problema que tiene dos componentes: una teoría *T* y una sentencia (del lenguaje de la teoría) que representa un hecho *E*. De aquí, mediante un proceso inferencial abductivo, se obtiene *C*. Entonces de *T* y de *C* se deducirá *E*. Esquemáticamente,

- (*T*, *E*) ABDUCEN *C*
- *T, C* DEDUCEN *E*

En su contexto DEDUCEN tiene un sentido específico, el cual puede cambiar de un contexto a otro. Sin embargo, como se indicó

anteriormente, si se trata de una relación de consecuencia de clausura, aunque cada contexto tenga sus características propias (debido, por ejemplo, a que se asuman distintos conjuntos de postulados, o algunos de estos sean muy específicos del campo de que se trate), se mantiene un estilo clásico, de manera que el modelo de explicación, en lo esencial, es el depurado a partir de la propuesta de Hempel y Openheim. La exigencia de que sea definible un cálculo lógico adecuado, es decir, que venga a ser correcto y completo, abunda en la supremacía del punto de vista clásico.

5. Observaciones finales

Para cada práctica científica podemos considerar una "lógica subyacente" como aquélla que codifica los modos de razonar en dicha práctica ¿Sólo se admiten conclusiones que han sido obtenidas, por así decir, siguiendo estrictamente los correlatos de los pasos marcados en los sistemas clásicos? Un examen del desarrollo de algunas disciplinas, como la inteligencia artificial, por ejemplo, viene a sugerir que ciertas conclusiones se obtienen siguiendo un sistema inferencial difícilmente modelado por los sistemas de lógica clásica. Casos como los de conclusiones en cuya obtención se ha manejado la negación por fallo, o cuando se ha barajado una información disponible, a todas luces incompleta y pendiente de una posterior ampliación, sugieren que la modelación de los intentos explicativos de carácter lógico debería tener en cuenta una noción de deducibilidad no constreñida a la clásica. Algunas relaciones estudiadas no verifican monotonía, por lo que no son relaciones de clausura y presentan el atractivo de constituir, a su vez, una herramienta explicativa de la génesis de ciertas hipótesis científicas.

Atendiendo a esa cierta dualidad entre abducción y deducción, si la primera es un modo de inferencia que permite obtener hipótesis explicativas ¿Por qué se ha trabajar necesariamente con la negación clásica, obteniendo abducciones mediante el recurso casi exclusivo de aplicar la contraposición? Es innegable la importancia lógica del "metateorema de la deducción", pero el camino de la obtención de abducciones tal vez sea más amplio que el marcado por el estilo de inferencia clásico. De acuerdo con el esquema planteado en la última sección, una vez que se ha abducido una hipótesis, ésta será

explicativa si junto con la teoría se "deduce" el hecho, en alguno de los sentidos que, como hemos visto, puede tener la noción de deducción. Así pues, a tenor del desarrollo de la lógica y una concepción de carácter menos fundamental de la misma, podríamos afirmar que el modelo nomológico-deductivo aparece como un marco general, especificable en tantos modelos concretos como maneras se cuenten de entender el requisito de la deducibilidad.

Referencias

Aliseda, A. (1997). *Seeking Explanations: Abduction in Logic, Philosophy of Science and Artificial Intelligence*. Institute of Logic, Language and Computation, ILLC Dissertation Series, 1997-4. Amsterdam.

Aliseda, A. (2006). *Abductive Reasoning. Logical Investigations into Discovery and Explanation*. Springer, Synthese Library/Volume 330. Dordrecht.

Amor, J. A. (2006). *Compacidad en lógica de primer orden y su relación con el teorema de completud*. Ediciones de la Facultad de Ciencias de la U. N. A. M. México

Eberle, R., Kaplan, D. & Montague, R. (1961). "Hempel and Oppenheim on Explanation", *Philosophy of Science*, vol. 28, 4, pp. 418-428

González, W. J. (coord.) (2002). *Diversidad de la explicación científica*. Ariel. Barcelona

Halonen, I., Hintikka, J. (2005). "Toward a Theory of the Process of Explanation", *Synthese*, vol. 143, 1-2. 5-61

Hempel, C. G.-Openheim, P. (1948). "Studies in the Logic of Explanation", *Philosophy of Science*, vol 15, 2, pp. 135-175

Hodges, W. (1984). "Elementary logic", en D. Gabbay & F. Guenthner (eds.): *Handbook of Philosophical Logi*. Reidel. Dordrecht, pp. 1-129

Kaneale, W. & M. (1974). *El desarrollo de la lógica*. Trad. J. Muguerza. Tecnos, Colección Estruntura y Función. Madrid

Makinson, D. (2005). *Bridges from Classical to Non Monotonic Logic*. College Publications. Londres

Mayer, M. C., Pirri, F. (1993). "First order abduction via tableau and sequent calculi", *The Bulletin of the IGPL*, vol. 1, pp-99-117

Nepomuceno, A. (2002). "Scientific Explanations and Modified Semantic Tableaux", en L. Magnani, N. J. Nersessian, C. Pizzi (eds.) *Logical and Computational Aspects of Model-Based Reasoning*. Kluwer Applied Series. Dordrecht, pp. 181-198

Nepomuceno, A., Soler, F. (2007). "Validez y contexto inferencial" (Pendiente edición Actas del XIV Congreso Internacional de Filosofía, Sociedad Mexicana de Filosofía, Mazatlán, 2007)

Peirce, Ch. S. (1958). *Collected Papers*, vol. 2, 1931-1958. Hartshorn et al. (eds.). Harvard University Press. Cambridge (Ma)

Schurz, G. (2005). "Explanations in Science and the Logic of Why-Questions: Discussion of the Halonen-Hintikka Approach and Alternative Proposal", *Synthese*, vol. 143, 1-2, pp. 149-178

Soler, F. (2005). Modelos formales de explicación en Lógica e Inteligencia Artificial Electronic version (doctoral dissertation). Editorial: Summa Logicae Siglo XXI, http://logicae.usal.es/

Velasco, D. (2006). "Modelos inferenciales". *La razón comunicada V. Materiales del Taller de Didáctica de la lógica*. México

A MOST ARTISTIC PACKAGE OF A JUMBLE OF IDEAS*

Fernando Ferreira
Universidade de Lisboa

O último artigo de Kurt Gödel deu à estampa em 1958 na revista suiça *dialectica*, num número especial dedicado ao septuagésimo aniversário do seu amigo e correspondente Paul Bernays. O presente sumário resume um ensaio publicado num número especial da *dialectica* dedicado ao cinquentenário deste trabalho de Gödel (vol. 62, pp. 205-222, 2008).

O artigo de Gödel, intitulado «*Über eine bisher noch nicht benützte Erweiterung des finiten Standpunktes*», apresenta uma extensão do ponto de vista finitista em relação à qual se pode demonstrar a consistência da aritmética de Peano. Como é bem sabido, em 1931 Gödel mostrara que o ponto de vista estritamente finitista não é suficiente para acomodar um tal argumento de consistência. A extensão proposta por Gödel introduz uma nova entidade abstracta, a saber: *funcional computável de tipo finito sobre os números naturais*. O argumento de Gödel baseia-se numa transformação sintáctica (hoje conhecida como interpretação *dialectica*) da aritmética de Peano num cálculo proposicional cujos termos se referem às novas entidades abstractas. Argumentamos no nosso ensaio que a introdução desta nova entidade ilustra, uma vez mais, a conhecida posição *realista* de Gödel e que está em sintonia com a sua visão da inexauribilidade da matemática (o que o leva a defender – em estudos fundacionais – a postulação de tipos cada vez mais elevados; p.ex., em teoria dos conjuntos, através de axiomas de existência de cardinais inacessíveis). Defendemos que a noção Gödeliana de *funcional computável de tipo finito sobre os números naturais* é uma noção primitiva, não redutível a noções da teoria dos conjuntos, a ser investigada através de novas intuições axiomáticas.

* O título deste trabalho (que em português se pode traduzir livremente por «Um pacote muitíssimo artístico de uma amálgama de ideias») é tirado da p. 110 do artigo de Georg Kreisel intitulado «Gödel' s excursions into intuitionistic logic» publicado em *Gödel Remembered*, Nápoles, Bibliopolis (1987).

Simultaneamente, discutimos o carácter *intensional* da nova noção. Tecnicamente, esboçamos a demonstração do teorema de William Howard (1973) que mostra que o axioma da extensionalidade não tem uma interpretação *dialectica*. Observa-se que o argumento de Howard é válido em circunstâncias muito gerais baseando-se, fundamentalmente, no facto de que todo o termo fechado do sistema de Gödel é *majorizável*. No ensaio também se discutem brevemente as hesitações intelectuais de Gödel que o levaram, durante oito anos, a estar indeciso sobre a eventual publicação de uma tradução inglesa do seu artigo. Passados estes oito anos, a tradução não se materializou.

A segunda parte do ensaio aborda aplicações recentes da interpretação *dialectica* à análise de demonstrações matemáticas (via a extracção de informação de natureza computacional). Estas aplicações estão baseadas numa modificação da interpretação de Gödel (que usa ainda as transformações sintácticas de Gödel), a denominada *interpretação funcional monótona* de Ulrich Kohlenbach (1993). Nesta interpretação, a noção de funcional majorizável desempenha o papel central. Descrevemos o programa de Teoria da Demonstração *aplicada* (em especial à Análise Funcional Numérica), conhecido por *Proof Mining*, comentando os seus sucessos e benefícios. Na parte final do ensaio, discutimos uma recente interpretação funcional, a *interpretação funcional limitada* do autor deste sumário em colaboração com Paulo Oliva (2005). Esta nova interpretação funcional baseia-se numa transformação sintáctica distinta da de Gödel e que «age como se» todo o funcional fosse majorizável. Mostramos como os resultados teóricos de Kohlenbach se podem ver, de uma forma muito perspícua, à luz desta nova interpretação.

A interpretação *dialectica* de Gödel envelheceu muito bem nestes cinquenta anos. Acreditamos que ainda não foi dada a última palavra sobre os seus méritos fundacionais e que as suas potencialidades (e as de interpretações semelhantes) ainda não estão cabalmente compreendidas.

LIMITATIONS OF THE PREDICATE CALCULUS. THE CASE OF DONKEY SENTENCES

María J. Frápolli
Departamento de Filosofía I
Universidad de Granada

1. El contexto general

That the first order predicate calculus has its shortcomings when applied to the complexities of natural languages is no news. This is the background in which extended and non-standard logics find their rationale. But besides adding new operators or rejecting some of the rules governing them, there are deeper ways of contesting the logical path that the predicate calculus has converted in the paradigm of logical reasoning. Showing that the complexities of natural languages will require finer and more flexible methods of analysis than those allowed by the standard predicate calculus will be one of the aims of this paper. Nothing new under the Sun. Logicians, linguists and philosophers in the last century has overtly defended or covertly suggested that a change would be desirable.

Our purpose here will be to illustrate the need of a new look at the logical features of natural languages focusing on a set of examples, the so-called "Donkey sentences", particularly reluctant to analysis. The trademark of donkey sentences is the co-existence of logical operators, i.e., connectives and quantifiers, together with anaphoric links among singular terms. The main difficulty for a suitable analysis, capable of being smoothly accepted by the varied group of theories that have risked a proposal, lies in the issue of the identification of the operators (and operations) involved and their relative scope.

We will use Donkey sentences to illustrate a general proposal about conditionals. Conditionals, as second-order binary operators, should be understood, we will maintain, as some kind of quantifiers. When there is no adverbial operator that modifies its force, conditionals are general quantifiers ranging over sets of situations. When adverbially modified, the quantifier might cease to be a general quantifier and turn into one of the non-standard weaker quantifiers,

"most", "a few", "seldom", and in the limit case of negation, it becomes an existential quantifier. Conditional sentences are thus quantified sentences, and the adverb "then" that usually lies implicit in the consequent clause works as an anaphoric pro-noun, although from a logical point of view we should say an anaphoric "pro-adverb", which is linked to its head, the general (conditional) quantifier. This way of looking at conditional logic, i.e., considering the conditional particle if as a universal quantifier with the meaning of "in every (some, almost all, etc.) situation" and interpreting the particle then as an anaphoric pro-adverb pointing to the situation at issue, allows to take a fresh look at the problem of Donkey sentences.

2. Donkey Sentences: A Study Case

Donkey sentences were introduced in the philosophical debate by Peter Geach in the sixties (Geach 1962: 155-156) and, since then, they have been a field in which every theory of quantifiers and pronouns in natural languages tests their merits. Although Donkey sentences are primarily used to illustrate (and explain) the mechanism of anaphora, we will not focus on this referential operation, but will be interested in the identification of the operators (connectives and quantifiers) in the logical form behind this kind of sentence.

Examples of Donkey sentences are:

(1) If Pedro has a donkey, he beats it,
(2) When a farmer owns a donkey, he beats it,
(3) Always that a farmer owns a donkey, he beats it now and then.

Examples (1) and (2) are the classical cases, and example (3) is a more complex version due to D. Lewis (1975).

One of the pervasive difficulties listed in the bibliography has been explaining why, for a question of scope, (1) cannot be translated as

(1.a) $\exists x$ [(Donkey (x) & Owns (Pedro, x)) $\rightarrow$ Beats (Pedro, x)]

This rendering is inappropriate because it would give priority to

the existential quantifier over the conditional, converting (1) in an existential claim, which it is not.

The alternative option (1.b),

(1.b) $\exists x$ (Donkey (x) & Owns (Pedro, x)) $\rightarrow$ Beats (Pedro, x),

i.e., giving the conditional wider scope, would leave the variable in the consequent clause unbound and thus convert it in an open formula, which it is not either,

To overcome this difficulty, it has been a common place to translate (1) using a general quantifier instead, as in (1.c),

(1.c) $\forall x$ [(Donkey (x) & Owns (Pedro, x)) $\rightarrow$ Beats (Pedro, x)].

Regarding (1.c), a point of discomfort has been how (and why) the expression "a donkey" is now translated using a general quantifier, whereas in narrower contexts, like (1.d),

(1.d) Pedro owns a donkey,

it would have been understood as involving an existential quantifier binding the variable that accompany the monadic first order predicate, as it happens in (1.e), the standard translation of (1.d),

(1.e) $\exists x$ [(Donkey (x) & Owns (Pedro, x)].

To illustrate the debate behind the difficulties listed above, it will be enough to recall the treatment of donkey sentences in three major contemporary theories: Dynamic predicate logic (DPL), Game theoretical semantics (GTS) and Discourse representation theory (DRT).

The ideas behind DPL and GTS are similar and consist in distinguishing two notions of scope (GTS) or of binding (DPL), which in this case amounts to the same, although the general differences between the two views are notorious. While DPL sticks to the Principle of Compositionality (and prides itself on it), GTS rejects the principle as a general *desideratum* of any semantic theory. The Principle of Compositionality cannot be maintained together with the

branching quantifiers that are characteristic of GTS. DPL and GTS also share their shortcomings, being both unable to explain the mechanisms behind anaphoric links when the head is either in a general or a negative sentence.

GTS denounces that the notion of scope that we use in our standard *calculi* embodies, in fact, two different notions that answer for different relations among the ingredients of propositions. One of them accounts for the relations of logical priority among logical operators, the other accounts for the binding scope of quantifiers. In complex cases like those of Donkey sentences, the two scopes do not go together and hence the translation problems into a *calculus* that, like standard first order *calculus*, does not make room for the distinction. With roughly the same range of problems in mind, DPL introduces two kinds of binding for variables. An occurrence of a variable in a formula can be free in the standard sense, if from a syntactical point of view it lies outside the scope of quantifiers in the formula, while anaphorically referring to some of these quantifiers that acts as the head of the anaphora. These variables would be, using Sandu' s terminology (Sandu 1997), syntactically free but semantically bound to their head. Bifurcation, in scope or binding respectively, is meant to differentiate the logical priority of operators, that in Donkey sentences like (1)-(3) correspond to the conditional over any quantifier, from the scope of quantifiers that have to reach all variables in the formula. DPL would render (1) as (1.b), and GTS would use its two-dimensions language to display the independence of both tasks.

An interpretation of a sentence in DRT is a set of referential markers and conditions. This set is called "a Discourse Representation Semantics" (DRS). Referential markers are singular terms, and conditions are formulae that the individuals referred to by the referential markers have to satisfy. A DRS for (1) would be something like the following:

(DRS.1) (a, b, donkey (b), beats (a, b)),

where "a" and "b" are referential markers, corresponding to "Pedro" and "a donkey", and the rest are conditions on them. Although DRT would interpret (1.d) as (1.e), it would treat (1) as if it had a structure similar to (1.c), with a universal quantifier. These

treatment (see, for instance, van Eijck & Kamp, 1997) needs to assume that the logical role, and then the operator behind, of indefinite descriptions, as "a donkey", vary depending on context, having sometimes the import of a universal quantifier, sometimes that of an existential quantifier.

Keeping track of what has been said so far, donkey sentences, while perfectly grammatical and admissible sentences, pose translation difficulties that reflect difficulties in understanding the logical ingredients of propositions expressed and the relations among them. These difficulties have required modifications of the standard notions of scope or of binding, or else making of indefinite descriptions context-dependent expressions in an unusual way.

3. Conditionals as General Quantifiers

The background picture behind the new look at Donkey sentences that we would like to defend here can be summarize, in a nutshell, in the idea that unmodified[1] conditionals are instances of universal quantifiers (*if* as an instance of $\forall$) and that quantifiers have, in natural languages, an intra-linguistic, non-representational, inferential, import. Making the first claim more precise, conditionals are, if they have to support an inferential interpretation, irreducible[2] n-ary devices, n >1. For the second claim we will not argue in this paper although it would be needed to bring into the mind a couple of the consequences of it. One of them is that we reject the common objectual interpretation of quantifiers, its restriction of only allowing nominal variables and, with it, the connection between quantification and ontology. Another one is that quantified sentences do not have primarily the role of representing reality; they rather display inferential connections from the speaker's point of view. Its role is not to be used to express propositions and thus they can be said to be true or false only in a derivative way. Some effects of these consequences will be patent in the interpretations of Donkey sentences in section 4

[1] By "unmodified" we understand conditionals in which no expressions occur that constraint the force of the universal quantifier. Some adverbs, like "usually", "often", "seldom" etc. indicate that the quantifier at issue is weaker than generality.
[2] In the sense explained by Keenan (1987).

below.

The picture that we favor has to be distinguished from two views on conditionals that might appear quite similar. One is the interpretation of conditionals as generalized quantifiers (Barwise and Cooper 1981). The other one is the deductive view of conditionals. Generalized quantifiers are relations among sets of individuals. And they are true when the specified relations among the sets involved hold. Our view understands conditionals as intra-linguistic devices that display permissions to infer propositions under particular circumstances. The deductive view on conditionals roughly upholds that, in conditionals, the information in the consequent follows, in a logically valid inference, from the information in the antecedent together with some added premises. This view is too strong and also takes a wrong perspective. It amounts to saying that a conditional is true when the consequent is true in all models in which the antecedent, together with the relevant premises, is true. It takes a model-theoretical, external, perspective. In our view a conditional is neither true nor false, they do not represent situations but rather have the import of authorizations of informational transfer contextually constrained.

In his (1984), van Benthem initiated a research program on conditional logic. Conditional logic is the logic of conditional statements. Van Benthem begins his paper with the following words: "Conditional statements occupy a central place in reasoning, and hence their proper analysis is a principal task of logic" (1984:303). Van Benthem´s main claim about conditional statements and its logic is that "if" works as a generalized quantifier, as a function connecting sets of antecedent and consequent "occasions", as he says. Notice that van Benthem's thesis is that "if" works as a generalized quantifier, i.e., as a determiner that expresses a relation between sets of individuals. What we want to put forward here is that *if* is a general (not necessarily unrestricted) quantifier, interpreted as an inferential device of transference of information. Backgrounds behind van Benthem's position and ours are quite divergent, although we are applying his insights to develop a stronger thesis. The stronger thesis is that, in natural languages, *if* and *all* perform the same task, being the difference between them one in the range of the "entities" quantified over. Conditionals are universal quantifiers ranging over sets of situations, circumstances, models or possible worlds ─ every one

chooses their favorite theory. Thus, *if* is an instance of *all* for the case of situations (or your favourite alternative). The accompanying particle "then" works, on van Benthem's view that we borrow, as an anaphoric pronoun bound by the universal quantifier. Van Benthem wants to keep *if* as a separate quantifier, with its own peculiarities, and others (Lapierre (1996), for instance) have followed this general path, characterizing *if* as one among the general class of Generalized Quantifiers. Lapierre (1996: 237-8), nonetheless, actually identifies what he calls a "conditional relation" with the quantifier "all". From the point of view of its formal properties, one might want to characterize a minimal conditional logic, and to continue then adding axioms to cover a wider range of determiners. Nevertheless, if the aim is the meaning of natural language expressions, all these subtleties are inapplicable. This thesis on natural languages does not diminish in the least the interest of van Benthem's project for formal semantics.

4. Back to Donkey sentences

With a view on conditionals as general quantifiers, the structure of Donkey sentences can be displayed in a systematic, straightforward way that also explains the difficulties, and other alternative ways of solution, like those of GTS, DPL, and DRT.

All these proposals agree in that the conditional in (1)-(3) has the logical priority over the quantifiers. And they are right. GTS signals it making the conditional independent of the quantifier. DPL establishes the priority of the conditional maintaining the variables in the consequent syntactically free (although semantically bound by the existential quantifiers) and DRT eliminates the existential quantifiers altogether. DRT is (accidentally) closer to our view in that it acknowledges the occurrence of a general quantifier. But, contrary to DRT' s explanation, the universal quantifier in (1.c) does not translate the indefinite noun phrase "a donkey" that, even in this context, has an existential quantifier at its logical form, but the conditional particle *if*, which is a universal quantifier on its own, ranging over sets of situations. This universal quantifier has indeed the logical priority in (1). Thus, those who have felt that there were a general quantifier with wider scope were right: the *if* quantifier. But also those who have felt that there must be an existential quantifier with narrower scope are

right, although related to a different logical ingredient. In fact, there are at least two quantifiers at work in the simplest Donkey sentence, the universal *if* quantifier which has the logical priority and the usual existential quantifier that falls under the scope of *if*. The quantifiers in (1) are thus (a) a universal quantifier translating the conditional particle *if*, meaning "in all situations s", and having the logical priority, and (b) an existential quantifier related to the indefinite description "a donkey". There also are three anaphoric links, one related to the pronoun "it" which is a variable bound by the existential quantifier that occurs in "a donkey", another related to the pronoun "he" which head is the proper name "Pedro", and still another related to the (implicit) *then* particle which also is a variable bound by the universal quantifier signaled by the *if* particle. Then a better translation of (1) into a semiformal language that discloses its logical structure would be (1.f),

> (1.f) $\forall s$ [$\exists x$ (In s, Donkey (x) & In s Owns (Pedro, x)); In s Beats (Pedro, x)].

Syntactically "in s" is an adverbial phrase, and from a logical point of view is a circumstance shifting operator.

Donkey sentences are much more complex than that, and it is no wonder that they have deserved the enormous amount of ink and paper already devoted to them. We have already listed the *if* quantifier, the existential quantifier, and the three anaphoric links marked respectively by "then", "he" and "it". The anaphoric links cannot be completely characterized as repeated bound variables[3], but require their own operator, what adds three more operators to the logical form of (1). The total sum will be five higher-order functions, which explain the difficulties enclosed in this kind of examples.

With this analysis in mind, let us go back to examples (2) and (3) and see how it behaves. Example (2) above,

[3] A promising way of treating anaphoric links is using a second order quantifier-like identity operator, as the one used by Williams (1989). An alternative, and equivalent, way is using Geach's reflexivity operator, (Geach 1962). A treatment of anaphora using a strategy close to Geach´s operator can be found in Böttner (1992), and he draws the idea back to Suppes (1974) and Keenan (1987).

(2) When a farmer owns a donkey, he beats it,

adds a new existential quantifier to the logical form of (1) corresponding to the indefinite description "a farmer". This quantifier also falls under the scope of the universal conditional operator. Thus, a translation of (2) into a semiformal language would be something like,

(2.a) ∀s [∃xy (In s, Donkey (x) & In s, Farmer (y) & In s, Owns (y, x)); In s, Beats (y, x)].

Example (3) poses an interesting challenge. Lewis uses it to discuss the range of the quantifiers involved, and says that in most interpretations, (3) will express a contradictory proposition due to the divergent meanings of "always" and "now and then". Independently of the weight we give to Lewis' conclusion, our proposed account of conditionals as quantifiers will explain why (3) is not contradictory, although it is somehow paradoxical in the sense that the adverb "always" does not add anything to the content. (3),

(3) Always that a farmer owns a donkey, he beats it now and then, is only a stylistic variant of (4),
(4) When a farmer owns a donkey, he beats it now and then.

In both cases, "always" and "when" signal the presence of a universal rule, and the users of (3) and (4) are displaying their beliefs about all situations of a certain kind. They are all those situations in which there is (at least) a farmer, (Fx), and (at least) a donkey, (Dy), and in them the farmer beats the donkey now and then. Whereas "always" or "when" encode sets of situations, "now and then" talk about instants of time. No contradiction then, although the contradiction obviously rises if, as happens in every standard analysis, "always" and "now and then" are interpreted as sharing their range. (4) would have a structure such as (4.a)

(4.a) ∀s [(∃x) (∃y) (In s Donkey (x) & In s Farmer (y) & In s Owns (x, y); ∃t (In s Beats (x, y, t))].

Our paraphrases do not belong to the predicate logic. Sub-formulas in (2.a), (1.f) and (4.a) are connected by a semicolon ";", which is not a connective of any standard calculus (and for which we have not given a definition). The reason lies outside the topic of the interpretation of conditionals as quantifiers, although it falls in the middle of the intra-linguistic, non-representational interpretation of quantifiers that, in our view, acts as general background.

Quantified sentences do not express truth bearers, because their role does not consist in representing states of affairs. They do not express propositions and cannot be fully translated in a linear, static notation. Among the varied formal proposals in the semantic arena, DRT and Situation Semantics are probably the most appropriate systems.

Donkey sentences also provide with examples in which apparently the Principle of Compositionality fails. From the point of view taken here, the reason of this failure is not the presence of complex anaphoric relations. The Principle of Compositionality simply does not apply, because what we reach after having processed a Donkey sentence is a complex network of inferential permissions and prohibitions. Exactly the same situation we encounter any time that there are quantifiers – in fact any kind of second order logical operators – at work. The playground of the Principle of Compositionality is the practice of expressing propositions. This is not the case when quantifiers are been used. At this point I remember Ramsey's words: "Many sentences express cognitive attitudes without being propositions; and the difference between saying yes or no to them is not the difference between saying yes or no to a proposition" (1931: 237-238). Conditional sentences are of this kind, they open restricted paths to transport information from some situations to others. They do not describe the world around but in their barest form show the speaker's inferential network.

In trouble-free examples, for uncomplicated semantic questions, the formal options that one chooses are, for most purposes, quite irrelevant, even if the philosophical backgrounds that support them are not subtle enough to undertake the task of analysing natural languages. But Donkey sentences are not interpreters-friendly. They resist any approach that does not incorporate the appropriate tools and perspectives. This proof-resistant nature is one of the sources of the success of Geach's examples, and a reliable guide for sharpening our

analytical equipment. Understanding conditionals as general quantifiers seems to pass the test. This is not its only merit, but we believe that it is at least a reason to attentively consider this proposal.

References

Barwise, J. (1989). *The Situation in Logic*. Stanford, Center for the Study of Language and Information.

Barwise , J. and Cooper, R. (1981). "Generalized quantifiers and natural languages", *Linguistic and Philosophy* 4, 159-219.

Barwise, J. and Etchemendy, J. (1987). *The Liar: An essay on Truth and Circularity*. New York, Oxford University Press.

Böttner, M. (1992). "Variable-free semantics for anaphora". *Journal of Philosophical Logic* 21, 375-390.

De Swart, H. (1996). "(In)definites and Genericity". In Kanazawa, M., Piñón, C. and de Swart, H. (1996), pp. 171-194.

Geach, P. (1962). *Reference and Generality. An Examination of Some Medieval and Modern Theories*. Ithaca and London: Cornell University Press.

Hintikka, J. and Sandu, G. (1997). "Game-theoretical semantics". In van Benthem and ter Meulen (eds.) (1997), pp. 361-415.

Kamp, H. and Reyle, U. (1993). *From Discourse to Logic. Introduction to Modeltheoretic Semantics of Natural Language, Formal Logic and Discourse Representation Theory. Part I.* Dordrecht, Kluwer Academic Publishers.

Kanazawa, M., Piñón, C. and de Swart, H. (1996). *Quantifiers, Deduction, and Contexts*. CSLI Publications.

Keenan, E. L. (1975). *Formal Semantics of Natural Languages*. Cambridge University Press.

Keenan, E. L. (1987). "Unreducible n-ary quantifiers". In P. Gärdenfors (ed.), *Generalized Quantifiers. Linguistic and Logical Approaches*. D. Reidel Publishing Company.

Lapierre, S. (1996). "Conditionals and Quantifiers". In van der Does and van Eijck (1996), pp. 237-254.

Lewis, C.I. and Langford, C. H. (1959). *Symbolic Logic*. New York, Dover.

Lewis, D. (1975). "Adverbs of quantification". In Keenan (1975).

Sandu, G. (1997). "On the theory of anaphora: dynamic predicate logic vs. game-theoretical semantics". *Linguistic and Philosophy* 20: 147-174.

Sandu, G. (2007). "Two Notions of Scope". In A. V. Pietarinen (ed.), *Game Theory and Linguistic Meaning*. Current Research in the Semantics/Pragmatics Interface, vol. xviii, Elsevier Science.

Suppes, P. (1974). "Model theoretic semantics for natural language". In Heidrich, C.H. (ed.): *Semantics and Communication*. Amsterdam, North Holland, 1974, pp. 284-313.

van Benthem, J. (1984). "Foundations of conditional logic". *Journal of Philosophical Logic*, 13, no. 3, 303-349.

van Benthem, J. and ter Meulen, A. (eds.) (1997). *Handbook of Logic and Language*. Cambridge, Mass., The MIT Press.

van der Does, J. and van Eijck, J. (eds.) (1996). Quantifiers, Logic, and Language. CSLI Publications.

van Eijck, J. and Kamp, H. (1997). "Representing discourse in context". In van Benthem and ter Meulen (eds.) (1997), pp. 179-237.

Williams, C.J.F. (1989). *What is Identity?*, Oxford University Press.

SISTEMAS EXPERTOS Y LINGÜÍSTICA COMPUTACIONAL: ESTRUCTURAS LÓGICO LINGÜÍSTICAS DE LENGUAS ANTIGUAS*

Cristina Barés Goméz

Abstract. This article discusses some of the problems regarding ancient semitic languages research when using computational tools. Computational Linguistics and Expert Systems implemented in hermeneutics methodology formalize a natural language better than classical methodologies.

1. Formalizaciones lógico-lingüísticas.

Los distintos avances en las lógicas matemáticas en los últimos años nos permiten acercarnos a diferentes formas de argumentaciones correctas. Las lógicas modales y la reciente lógica difusa (Trillas 1995) permiten abarcar un mayor espectro de las apreciaciones formales, pues ya no nos reducimos a los únicos valores de verdadero y falso. Ello nos permite continuar las diferentes investigaciones acercándonos cada vez más a la determinación formal del pensamiento mediante la resolución de problemas en ámbitos determinados. No podemos optar por una formalización completa que se asemeje al lenguaje natural, como ya demostró Gödel, pero mediante el estudio de los diferentes aspectos situados en problemas concretos podemos "hacer camino al andar".

* Esta investigación introduce parte de nuestro trabajo de Tesis Doctoral con beca I3P del Consejo Superior de Investigaciones Científicas (CSIC). Parte de este artículo expone parte de las investigaciones del proyecto PB-930107 dirigido hasta (el año) por J.L. Cunchillos. Actualmente las investigaciones de este proyecto han derivado en proyectos como el proyecto Bancos de Datos Semíticos Noroccidentales: Desarrollo y aplicación de nuevas tecnologías para el estudio y conservación de la documentación semítico-noroccidental del II y I milenio a. C. (HUM2007-65317, dirigido por J.P. Vita, CSIC-IEIOP), financiado por el Ministerio de Educación y Ciencia (España) en el marco del "Plan Nacional de Investigación Científica, Desarrollo e Innovación Tecnológica (I + D + I)", así como por la Unión Europea (Fondos Feder), del que también es resultado el trabajo que presentamos.

La determinación de una metodología progresiva que se vaya demostrando y se vaya probando es la esencia del método científico. Este método se plasma en las investigaciones lógico-lingüísticas gracias a la aportación de la informática. La técnica aporta el valor heurístico a las abstracciones lógico-lingüísticas para poder ir resolviendo los problemas y comprobarlos. Nos acercamos al razonamiento desde el punto de vista abstracto y la comprobación matemática. Además gracias a los avances en las nuevas tecnologías hemos pasado a la fase física.

Un sistema de ingeniería responde parcialmente a los problemas, cuyos límites residen en el propio sistema. Así hablamos del concepto de simulación, pues el objeto artificial se encarga de simular al objeto real, imita al real produciendo los mismos efectos. La simulación es la técnica para predecir el comportamiento de los sistemas en las computadoras digitales. Hay dos ámbitos de aplicación o estudio, el abstracto y el empírico. Los sistemas de ingeniería informática aúnan estas dos características como simulación del pensamiento para la comprobación de las teorías. Tienen un claro componente abstracto de teorías lógico-matemáticas. Es bien sabido que los avances en la programación derivan de avances en las teorías lógico-matemáticas. Las primeras programaciones eran en lo que se denomina lenguajes máquina, es decir en sistemas binarios de lógica bivalente. Ahora podemos hacer programaciones en lógicas sentenciales, pues tenemos los programas ensambladores que pasan las secuencias a lógica binaria. Así surgieron programas como Prolog o C+. Ello se produjo gracias a la llamada regla de resolución que permite implementar reglas de primer orden sobre el computador. Tras esta, Colmenauer en 1972 definió el programa Prolog que permite representar e interpretar el conocimiento mediante reglas clausulares o las llamadas cláusulas de Horn.

Por otro lado tenemos el componente físico de los ordenadores. Es la ciencia empírica de las computadoras, sistemas que son capaces de copiar, codificar y mover símbolos. Simon (1969) nos habla de una posible historia natural debido a su similitud con la naturaleza humana de sistemas nerviosos. Se habla de una puerta para entender el comportamiento humano a través del comportamiento del ordenador. Este es una simulación del otro. Así hablamos de la familia de artefactos llamados sistemas físicos de símbolos, los ordenadores y la mente humana. Sin embargo, igual que la escritura no es una copia

exacta del lenguaje, tampoco lo es la computadora del cerebro. El objeto artificial es una simulación. No hablamos de una imagen perfecta, pues entonces ya existiría la inteligencia artificial, sino de aproximaciones a los distintos procesos que realiza la inteligencia humana. Esta aproximación la vamos a hacer mediante el lenguaje humano.

Agrupamos así los diferentes avances en la Ingeniería del conocimiento desde dos paradigmas. El primero el paradigma simbólico y el segundo el paradigma biológico. El paradigma simbólico es el que vamos a desarrollar brevemente en cuanto compete a las estructuras lógicas, sistemas de argumentación que determinan la implementación posterior en los ordenadores. Sin embargo no podemos obviar el biológico, pues todo depende del tipo de software y hardware. La vertiente práctica de comprobación de hipótesis reside en el paradigma biológico.

2. Sistemas Expertos en Inteligencia Artificial.

Se considera que la fecha exacta de la creación de la Inteligencia Artificial fue 1956. En esta fecha se celebró un congreso en el Dartmouth College, New Hamspshire dirigido por McCartthy, Minsky, Rochester y Shannon (Álvarez Munárriz 1994). Su objetivo eran las investigaciones que toman como punto de partida la analogía del ordenador como forma de simular el comportamiento de la inteligencia humana.

Tras un periodo inicial y otro de consolidación surgen los llamados Sistemas Expertos. Su desarrollo es el objetivo de esta intervención. Estos sistemas son programas de ordenador cuyo comportamiento no difiere del comportamiento que tendría una persona experta a la hora de resolver un problema específico.

Sin embargo la forma de resolver problemas no sólo compete a la estructura argumentativa simbólica, datos + algoritmos. La dificultad reside en la formulación de las reacciones de un experto ante un problema concreto. El experto no sólo tiene en cuenta los determinados algoritmos. Sus procedimientos son heurísticos. Aquí el contexto, la situación, lo que normalmente se llama "sentido común", es lo que hay que formular. No sólo son argumentaciones cerradas. Tenemos la posibilidad de una estructura abierta que pueda abarcar las

distintas contextualizaciones necesarias para la resolución de cada uno de los problemas.

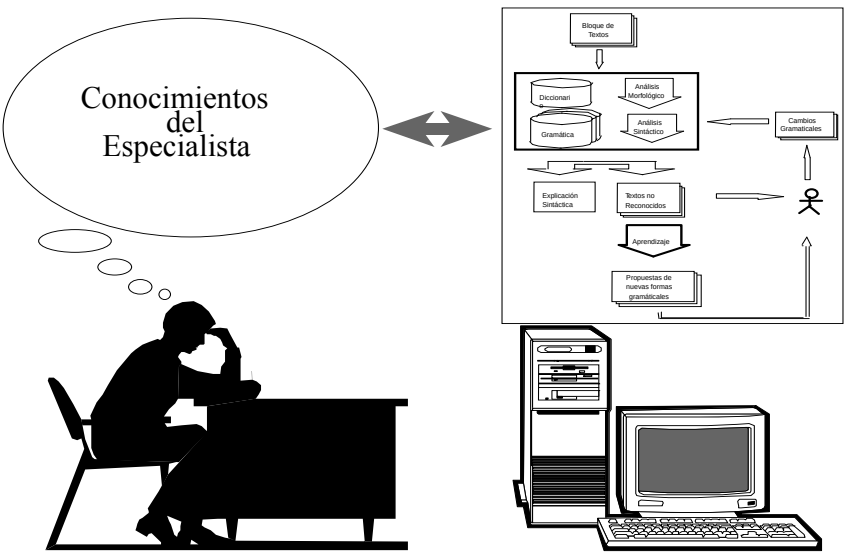
Según la definición clásica de un Sistema Experto podemos diferenciar varias partes en su estructura que determinan la arquitectura. Constan de tres subsistemas bien diferenciados. Aunque ello no impide la creación de diferentes módulos o niveles:

> - Una base de conocimientos. Es la base que en Ingeniería del conocimiento codifica y formaliza. Adquiere el saber compartimentado que posee un Experto en su materia. Sería un conocimiento declarativo, pero este al ser compartimentado y jerarquizado se va complicando. En un principio son las bases de datos simples. Sin embargo una base de datos no pasa de la acumulación de la información. Esto ya no es una acumulación, sino que mediante el programa se van a desarrollar diferentes procesos. La base es entonces un banco de conocimientos. No sólo son los datos, sino sus relaciones y las inferencias que hace un experto para llegar a sus conclusiones. No es un simple almacén. La forma lógica usada para la creación de estos bancos de conocimientos son los predicados y las relaciones lógicas entre predicados, enlazados por operadores lógicos. En lo que pretendo presentar, en concreto, para la representación gráfica, se usan sistemas de entidad relación[1]. Debido a que toda la estructura se basa en diferentes módulos, como iremos viendo, la posibilidad de utilizar lógicas más cercanas al razonamiento cotidiano como la lógica borrosa abre un campo de posibilidades a explotar.
> - Sistema de inferencias. Es el subsistema lógico que activa, controla y ejecuta el banco de conocimientos. Se implementa sobre la base de conocimientos. Permite hallar diferentes reglas según se vaya avanzando en la estructura del sistema. Aquí el campo de investigación serían las llamadas lógicas probabilísticas o también sistemas de razonamiento no monótonos. Estos sistemas se acercarían más a lo que llamamos "sentido común" (Nepomuceno 2006). Actualmente se han utilizado sistemas de jerarquías y conjuntos de acciones y

[1] Representación gráfica de los conjuntos usada frecuentemente en programación con bases de datos relacionales. Ver Cervigón (2007). Siabra (2007).

subacciones llamados ontologías computacionales. Todas estas herramientas pretenden acercarse a los distintos contextos que determinan los razonamientos. En cuanto al lenguaje, hay que tener en cuenta las teorías pragmáticas de Grice o los actos de habla de Searle. Las técnicas de representación y de razonamiento pueden ser muy variadas. Hay que centrarse en cada problema concreto. Se busca la solución que más se acerque a la resolución del problema por el experto.

- Interfaz de usuario. Es el dispositivo que se encarga de la comunicación entre el programa y el usuario. Presenta y facilita el trato de la información. Dependiendo del área de conocimiento que tratemos ha de ser uno u otro. Los expertos en los cuales se basan para la creación de los programas tienen también sus prioridades para un mejor manejo de las herramientas computacionales. Ello nos lleva a la búsqueda de una simbiosis entre los intereses de los distintos expertos en las distintas áreas y los informáticos que implementen los programas. Han de ser accesibles y útiles para ambos.



Proyecto Hermeneumática.

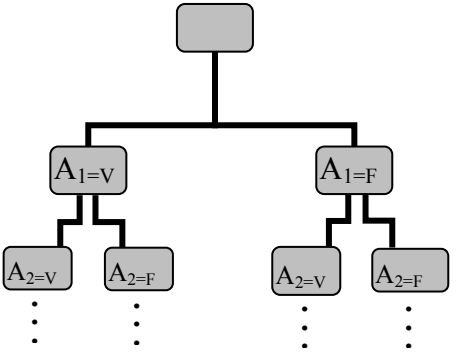
Para realizar todo esto se siguieron dos líneas principales (Cuena 2004):

- Deducción automática.

Se realiza una representación básica mediante cláusulas de Horn, programación lógica en primer orden. Esta línea formaliza las argumentaciones correctas que da un experto siguiendo sus pasos en la resolución de los problemas. Tendríamos entonces una base de conocimientos declarativa del tipo P_1, P_2,…, P_n que nos daría un resultado Q. Lo que habría que ver son los pasos a dar para poder decir si ese resultado es verdadero o no.

1. $P_1, P_2,\ldots, P_n \rightarrow Q$

Ello deriva que si en una base de conocimientos en lógica de primer orden se puede interpretar que Q es verdadera. Los caminos de la argumentación en deducción automática nos llevarían a realizar la transformación de Skolem, cuantificando universalmente las cláusulas. Mediante la interpretación del universo de Herbrand nos llevaría a describir un árbol semántico asignando significado a cada uno de los literales básicos. Daríamos una interpretación I a cada uno de los literales. Como tenemos una sentencia disyuntiva sería verdadera si alguna de sus ramas fueran verdaderas. Aunque las dos funciones no siempre son iguales lógicamente (la transformación de Skolem), se interpreta como tal. Así tenemos un árbol con ramas disyuntivas. Si una funciona, será satisfaccible. Si las cláusulas seguidas de la evaluación de la insatisfacibilidad dan como resultado el positivo, el proceso termina, si no, se vuelve a generar otro conjunto.

2. $P_1 \& P_2 \& \ldots \& P_n \& \neg Q$
3. $Vx_1, Vx_2,\ldots, Vx_m[C_1 \& C_2 \& \ldots \& C_n]$
4. $A_1 v A_2 v \ldots v A_n$
5. 

Hay otros modelos que computacionalmente son menos onerosos. Sin embargo, requieren almacenar la información sobre la rama en curso de proceso y si esta es infinita puede que el proceso no llegue al resto de las ramas. En diseños de modelos cognitivos se usan procedimientos heurísticos. El demostrador puede ser generado de diferente manera según la clase de problemas que se pretenda resolver.

- Modelización cognitiva derivada de la psicología y la filosofía cognitiva (Cuena 1995).

Esta línea fue planteada en un principio por Quilliam en el 68. Proponía una estructuración similar a la estructuración de la mente humana. Así se creaban diferentes redes cuyos nodos eran determinados conceptos. Los arcos eran relaciones entre los nodos, es decir relaciones entre los conceptos. Estas son las llamadas redes semánticas. Representan el conocimiento mediante complicados universos conceptuales. A partir de aquí Minsky crea el concepto de marco. Los marcos (frames) definen una serie de variables atributo. Caracterizan al concepto mediante atributos y toman determinados valores en dominios predefinidos.

Las jerarquías de conceptos o estructura prototípica de marcos hacen que en determinados casos tengamos que utilizar programaciones en lógica no monótona. Esta implicación sería más cercana al razonamiento de "sentido común".

1. $K_1 \rightarrow K_2$
2. $K_1, K_3 \not\Rightarrow K_2$

Todo ello dio lugar a los llamados Sistemas Expertos de 1ª generación. Pero se produjo un "cuello de botella" en la adquisición del conocimiento. No se podían caracterizar conocimientos complejos. Tenían reglas muy elementales y había que recorrer un camino muy largo para la formulación de las reglas complejas que usaban los expertos que servían de modelo. Tras estos sistemas de 1ª generación se fueron creando otros con programaciones más avanzadas. Se organizó el conocimiento en base a tareas específicas según las clases de problemas planteados. Así surgieron las diferentes divisiones en módulos para la resolución de determinados problemas que luego se unirían en una estructura total.

La creación de sistemas expertos nos facilita la formación de una estructura lógica implementada. Es una estructura abierta pues se va creando mediante la resolución de problemas concretos. Podemos además utilizar diferentes tipos de lógica para la creación de los diferentes tipos de módulos según se vaya necesitando. Las jerarquías son también jerarquías de inferencias.

Nos vamos a centrar en la formulación de un tipo de sistema experto de segunda generación basado en unidades cognitivas y resolución de tareas básicas. El Knowledge Structure Manager (KSM), fue propuesto y desarrollado por Cuena (2004) y aplicado al lenguaje Natural. Es un sistema basado en reglas simples de inferencia que son los primitivos de otras reglas de nivel superior. Se realiza mediante la compartimentación en diferentes módulos. Tenemos el KSM aplicado a la lengua natural. Hay que tener en cuenta la Ingeniería del Conocimiento para la creación de programas para la resolución de problemas concretos. Pero también hay que acercarse a las investigaciones en el campo del Procesamiento del Lenguaje Natural o Lingüística Computacional.

3. Lingüística computacional.

Los sistemas de Procesamiento del Lenguaje Natural (PNL) se basan en la implementación de las estructuras lógico-lingüísticas del lenguaje natural. Son usados en lo que se ha llamado Ingenierías de la Lengua. Son sistemas de traducción automática, reconocimiento de voz o búsquedas de respuestas. Estos sistemas han derivado de la formalización lógica del lenguaje y de las llamadas gramáticas formales. Se han implementado en sistemas de programación lógica, sobre todo en lógica de predicados de primer orden. Normalmente se hace una diferenciación básica entre sintaxis y semántica atendiendo a la diferenciación típica de la lógica clásica. Sin embargo, para formalizar las lenguas naturales es necesario tener en cuenta otros niveles.

Así la semántica lógica y la semántica del lenguaje natural son diferentes ámbitos a tener en cuenta a la hora de pasar a informática una lengua (Lavid 2005). Desde las gramáticas computacionales generativas de Chomsky (Sandoval 1999), ha habido estudios que dan mayor importancia a la morfología de las lenguas. Los primeros

estudios se derivaron de la aplicación computacional a lenguas con una morfología más rica que la del inglés. Es el caso del finlandés. Nosotros tratamos lenguas semíticas y estas se caracterizan por una morfología muy compleja. Los modos de tratar la morfología han sido sobre todo dos:

- Modelo en dos niveles PC-KIMMO. Fue propuesto por K. Koskenniemi (1983) para el finlandés. Se construye a partir de un lexicón en forma de árbol de letras. Utiliza autómatas secuenciales de estados finitos que trabajan en paralelo.

- Ejemplo 1:
Nivel lexico: "*talo*".
Nivel superficial: "*talo*".
- Ejemplo 2:
Nivel léxico: "*talo*".
Nivel superficial: "*talona*"

Ello se debe a que "talo" es casa en nominativo singular. "*talona*" sería como una casa. Habría sublexicones que son autómatas. Recogen los morfemas de una determinada clase. Uno de ellos analiza los caracteres superficiales y el otro los contrasta con la cadena donde se recoge la representación léxica de los morfemas. Tenemos que las reglas de combinación de los morfemas se especifican en el lexicón y no en reglas. Tenemos algo predeterminado por quien realiza el programa. La propuesta de este proyecto es basarse en la lengua misma y no en los conocimientos predeterminados del especialista. Todo ha de comprobarse sin dar nada por supuesto.

El PC-KIMMO atiende a una morfología concatenativa[2]. Para las lenguas semíticas se ha usado distintos niveles mediante el encadenamiento de transductores lo que ya no es el modelo inicial. Otro problema reside en la falta de atención del nivel sintáctico que se encuentra en las entradas léxicas. Ello provoca la necesidad de ampliación del léxico.

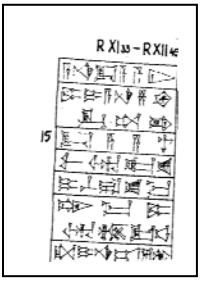
[2] Para las lenguas semíticas una morfología concatenativa resulta insuficiente, hay que tener en cuenta las relaciones incluso con elementos insertados en dos relacionados.

Un ejemplo de este sistema para lenguas semíticas es el caso del acadio[3]. Esta lengua es una lengua semítico-nororiental[4]. Se habló en la zona de Mesopotamia desde antes del III milenio y se conserva como lengua de erudición hasta la era cristiana. Normalmente se escribió con una escritura silábica cuneiforme usando ideogramas sumerios. Ello nos permite tener su vocalización. Como características fundamentales[5] podemos señalar el léxico triconsonántico fundamental, la declinación nominal se marca por vocales causales y diferentes esquemas radicales representan en el sistema verbal los tiempos, voces y modos. El problema de la morfo-fonología es que no sólo se determina mediante sufijos y prefijos, sino que además se realiza mediante infijos. Esta característica es resuelta en el parser con la creación de dos sublexicones. Uno de ellos se encarga de las raíces de las palabras y el otro de los prefijos, los elementos flexionados y los sufijos. Un parsers[6] para el acadio con este sistema nos daría para el término "*iparras*":

> Representación léxica: *ipar-os.*
> Representación superficial: *iparras*
> ***iparras*** Stem :p-r-s Parse[G, Durative, 3, c, s] Normalized[i, p, a, r, r, a, s]

Es una forma en inacabado del verbo "*parâsu*" (infinitive) que significa "cortar". El verbo estaría en 3ª singular y sería "corta" o "cortará" (él). Sin embargo podemos encontrar formas con sufijos que ya el parser no reconocería:





Código de Hammurapi. Siglo XVIII a. C.

Texto 168. Línea 17 (Bergmann 1953):
": 3ª singular inacabado.

*-*u*-*: infijo que puede ser subjuntivo o plural. Aquí es plural.

**-ma*: conjunción copulativa "y".
El verbo "*parâsu*" también puede ser "decidir". Aquí es "decidan". (Que los jueces decidan y…).

[3] Kataja y Koskenniemi (1988).
[4] Anexo. Lámina 2.
[5] Malbran-Labat y Vita (2005).
[6] http://www.wiglaf.org/akkadian/indexnew.html

- El modelo de morfología basada en la unificación y rasgos (Sandoval 1999). Son gramáticas de Chomsky independientes de contexto aumentadas con rasgos. Se utiliza el mismo tipo de formalismo para tratar todo tipo de unidades desde el morfema al discurso. Tras la teorías gramaticales de Chomsky surgieron gramáticas como la de unificación. Estas gramáticas no diferencian una estructura profunda de una superficial. No son necesarias las transformaciones para relacionarlas. Todos los elementos primitivos son tratados como rasgos. Ya no tenemos elementos como los fonemas o los morfemas, sino que todos son tratados como rasgos. La forma de unirlos es mediante la operación de unificación. El problema reside en que la información de un nivel no pasa a otro. Son también concatenativas y difíciles de manejar con un léxico amplio.

Estos dos modelos de herramientas computacionales, los dos niveles y la unificación mediante rasgos, no tratan muy bien el contexto paradigmático y las irregularidades. Estas están normalmente incluidas en el diccionario. Se necesitan léxicos muy amplios.

4. Sistemas expertos y lingüística computacional.

Vamos ahora a centrarnos en una lengua semítica, como hemos dicho antes, el ugarítico[7]. El ugarítico es una lengua que se habló en la región Sirio-Palestina en la ciudad de Ugarit en el II milenio a.C.[8]. Es una lengua de la familia de lenguas semítico-noroccidentales[9]. Se caracteriza por su rica morfología. Como lengua semítica su léxico viene determinado fundamentalmente por una raíz triconsonántica. Sin embargo, como suele ocurrir en estas lenguas, las raíces a veces cambian. Hay consonantes débiles que se asimilan con vocales y a veces las raíces son de dos consonantes. El ugarítico fue escrito con una escritura alfabética cuneiforme. En una lengua semítica no se escriben las vocales, sin embargo en el ugarítico tenemos algunas. Ello se debe a que las vocales “a”, “i” y “u” seguidas de la consonante débil “‘ ”(alif) si se escribían. Tenemos entonces “‘ a”, “‘ i” y “‘ u”[10].

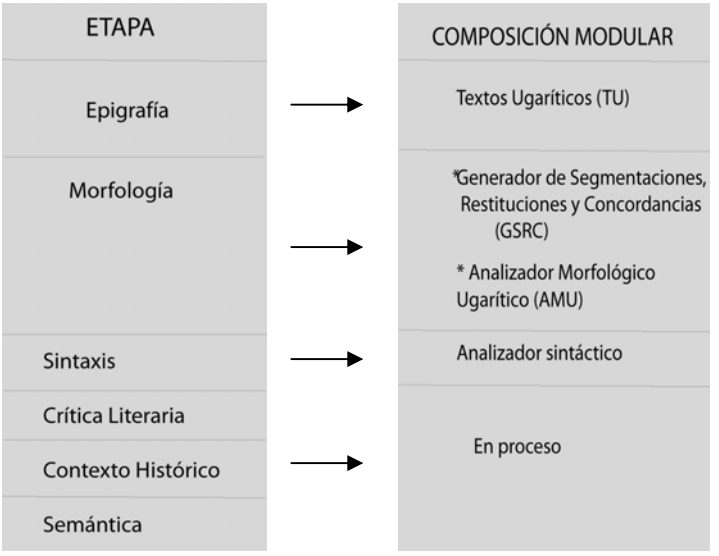
[7] Cunchillos y Zamora (1995). Segert (1984).
[8] Anexo. Lámina1.
[9] Anexo. Lámina 2.
[10] Anexo. Lámina 3.

Tenemos la vocalización original. Podríamos basarnos en el hebreo bíblico, pero los textos son copias unos de otros y la vocalización es muy posterior. Esta se debe a los estudios masorétiocos posteriores. Todo ello hace que el ugarítico sea una lengua, dentro del conjunto de las lenguas semíticas, cuyo interés fue en aumento.

En la Hermeneumática se realiza una mezcla entre los sistemas expertos y la lingüística computacional. Se trata una lengua en desuso del II milenio a. C. Fue una metodología creada y desarrollada por Cunchillos (2000). Su nombre deriva de la unión entre informática y hermenéutica. Tenemos pues una lengua de corpus, un corpus cerrado de textos, que se pueden ampliar según vayan apareciendo en los nuevos descubrimientos arqueológicos.

Al tratarse de una lengua en desuso no podemos comprobar nuestro conocimiento de ella mediante los hablantes. Partimos únicamente de los textos y para analizar, no sólo la lengua, sino la cultura que rodea a este pueblo del Antiguo Oriente Próximo, la informática nos da la posibilidad de comprobación. Mediante Sistemas Expertos podemos realizar una estructura compartimentada en distintos módulos con una lógica propia para cada uno de ellos. Además estos módulos se van a complementar unos con otros siguiendo los pasos de la lingüística computacional. Pero estamos ante una lengua que todavía plantea multitud de incógnitas. No escribe todas las vocales y estas son las que determinan la declinación y los modos. La fonética se tiene que conocer mediante lingüística comparada. La creación de los diferentes programas de las distintas lenguas del entorno puestas en relación, sería una puerta a la investigación que facilitaría el trabajo de muchos investigadores.

El Ugaritic Data Bank (UDB) es el programa creado por el equipo de Cunchillos que desarrolla los conceptos que hemos mencionado. Su composición es la siguiente:



A. El módulo informático I (TU). Se pasa del objeto, tablilla o inscripción a la línea de texto con accidentes epigráficos. Después se obtiene la Cadena Grafemática (CG) una vez limpia, restituida o segmentada. Algunas CG pueden ser PMD (Palabras en Morfología Desplegada), pero otras están incompletas. Así también tenemos CGI (Cadenas Grafemáticas Incompletas). Para resolver esto estará el siguiente módulo.

1. Objeto inscrito, (soporte e inscripción).
2. Línea de texto. Se adopta el texto base y se almacenan físicamente las líneas de texto con sus diferentes lecturas, colaciones.

La Cadena Grafemática es el resultado de la línea de texto transliterada y limpia de accidentes epigráficos. No es todavía una palabra.

B. El módulo II es el Generador de Segmentaciones, Restituciones y Concordancias (GSRC). Se encarga de preparar el texto para el módulo tercero, el morfológico AMU. Los procesos que transforman la línea de texto en PMD son: limpieza del texto, segmentación, división y unión de cadenas y restitución de cadenas

incompletas. La Palabra en Morfología Desplegada se caracteriza porque ya tiene significado y se forma por una serie de grafemas sin accidentes epigráficos y con todos sus constitutivos. Es el resultado del módulo II.

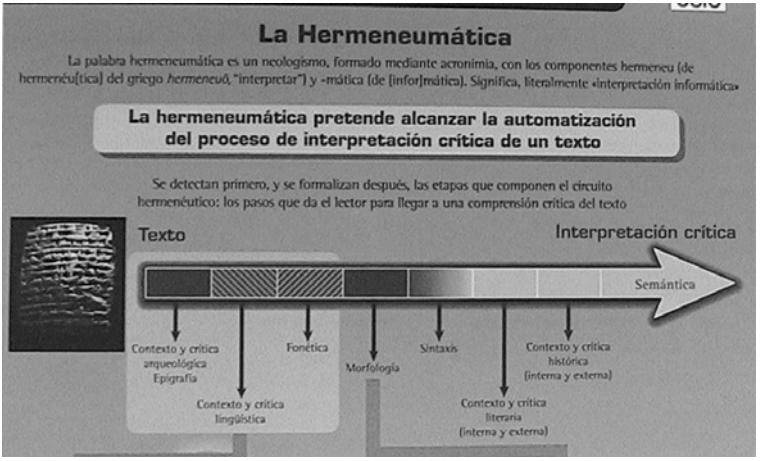
1. Tratamiento de Cadenas Grafemáticas.
2. La división y unión de Cadenas Grafemáticas. Cadenas Quebradas.
3. Concordancia de Cadenas Grafemáticas.
4. La restitución de Cadenas Incompletas.

C. El módulo III. Analizador morfológico Ugarítico (AMU):

Nivel 1: La PMD sin modificadores. AMU accede al léxico y la PMD coincide con la raíz/ radical.
Nivel 2: La PMD con modificadores morfológicos. AMU accede al léxico para que la raíz/radical coincida con el resto de cadenas grafemáticas.
Nivel 3: La PMD proviene de una raíz débil.
Nivel 4: Se describe la formación teórica de sustantivos que parten de raíces verbales y otros casos parecidos.

AMU aplica todas las reglas de descomposición a cada PMD. Las reglas identifican un número posible de modificadores morfológicos, algunos pueden ser erróneos. También aparecen posibles modificaciones susceptibles de cambios fonéticos al aplicarle las reglas fonéticas. Se aplican a su vez posibles cambios fonéticos con repercusiones gráficas, también las reglas restrictivas de comportamiento tanto de los verbos débiles como de los verbos con alif.

Pese a que AMU puede analizar cualquier palabra, sólo guardará los análisis de las PMD atestiguadas. Con los análisis almacenados se obtiene la Concordancia de Raíces Ugaríticas (CRU).



Proyecto Hermeneumática J. L. Cunchillos. (CSIC).

Mi línea de investigación se enmarca en este contexto. La función es determinar la estructura de una semántica formal con una metodología interpretativa. Analizar y exponer las distintas estructuras lógicas en los distintos módulos que den la aquitectura global de la producción del lenguaje natural. Tenemos pues una formalización de la lengua, pero con una metodología obtenida de la ingeniería del conocimiento y los estudios de la hermenéutica bíblica (Schökel 1986), siguiendo los métodos del método histórico crítico. Las aspiraciones de las humanidades en la búsqueda de una comprobación empírica mediante el método científico, se realizan gracias a las nuevas tecnologías. El ordenador formaliza la estructura lógica y comprueba los datos introducidos al ser implementados en reglas lógicas. Aquí reside la doble vertiente de la ingeniería del conocimiento, lo abstracto y lo empírico teniendo en cuenta todos los pasos de las distintas disciplinas que aportan cada una su granito de arena. La estructura viene dada por la lógica, la lingüística y su implementación informática. La metodología es una metodología hermenéutica. No podemos olvidar los distintos pasos que hay que dar para que la estructura de la lengua se plasme. Hay que pasar por la arqueología, la epigrafía, la fonética, la morfología, la sintaxis, la literatura y la historia. Es una metodología de hipótesis y comprobación, pues si el experto no introduce bien los pasos el ordenador no reconocerá el razonamiento y habrá que empezar de

nuevo. Además tenemos que se hace desde la lengua misma, intentando objetivar lo máximo posible sin introducir las categorías predeterminadas que podamos tener respecto sobre estas culturas con una distancia temporal tan amplia. Por último, es un sistema abierto, pues no estamos ante una traducción automática sino ante una automatización del proceso de interpretación de textos antiguos semítico-noroccidentales.

Bibliografía

Aaron Macks. http://www.wiglaf.org/akkadian/indexnew.html 2001-2002.

Alonso Schökel, L. (1986). *Hermenéutica de la palabra I. Hermenéutica bíblica.* Madrid.

Álvarez Munárriz, L. (1994). Fundamentos de inteligencia artificial. Murcia.

Bergmann, E. (1953). Codex Hammurabi. Roma.

Cervigón, R. (2007). *AMU. Analizador Morfológico Ugarítico*. En : *Las aguas primigenias. El Próximo Oriente Antiguo como fuente de civilización*.(Eds) Justel, J. J., Solans, B. E., Vita, J. P. y Zamora, J. A. Zaragoza.

Cuena, J. (1995). *La inteligencia artificial como simulación de procesos de razonamiento con conocimiento explícito*. ARBOR. Julio.

Cuena, J. (1998-99). *Sistemas inteligentes. Conceptos, Técnicas y Métodos de Construcción*. Madrid.

Cuena, J. y Molina, M. (2004).*Using Knowledge Modelling Tools for Agent-Based Systems: The experience of KSM*. En: *Knowledge Engineering and Agent Technology*. (Eds) Cuena, J., Demazeau, Y., García Serrano, A. y Treur, J. Ámsterdam.

Cunchillos, J-L y Zamora, J-A. (1995). *Gramática Ugarítica Elemental*. Madrid.

Cunchillos, J-L y Vita, J-P. (1998). *Introducción lectura crítica de documentos de II y I milenio*. Madrid.

Cunchillos, J-L. (2000). *Hermeneumática.* Madrid.

Grishman, R. (1991). *Introducción a la lingüística computacional*. Madrid.

Kataja, Laura and K. Koskenniemi (1988). *Finite – state description of semitic morphology: case study in Ancient Akkadian. En: Proceedings of the 12th International Conference on Computational Linguistics / organized by the John von Neumann Society for Computing Sciences ... [et al.]* Budapest : John von Neumann Society for Computing Sciences, [1988].

Koskenniemi, K. (1983). *Two-Level Morphology: A General Computational Model for Word-Form Recognition and Production*. Helsinki.

Lavid, J. (2005). *Lenguaje y nuevas tecnologías*.. Madrid.

Malbran-Labat, Fl. Y Vita, J.P. (2005). *Manual de lengua acadia.* Zaragoza.

Moreno Sandoval, A. (1999). *Lingüística computacional*. Madrid.

Nepomuceno, A. (2006). "Razonar con expresiones ambíguas". En V Congreso Sociedad de Lógica, Metodología y Filosofía de la Ciencia en España. Granada.

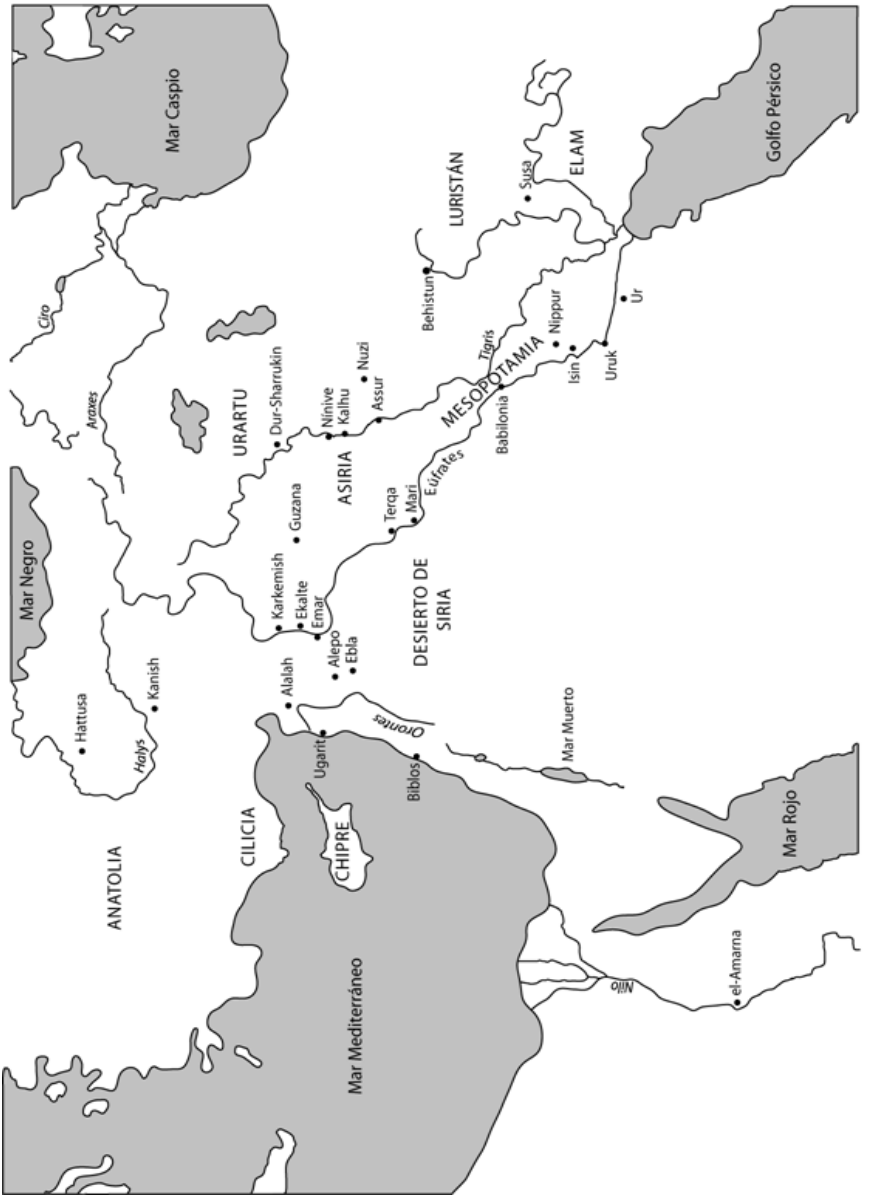
Segert, S. (1984). A basic grammar of the Ugaritic language with Selected Text and Glossary. Berkeley, Los Angeles and London.

Siabra, J. (2007). *El modulo sintáctico del Ugaritic Data Bank (UDB)*. En : *Las aguas primigenias. El Próximo Oriente Antiguo como fuente de civilización.*(Eds) Justel,J.J., Solans, B.E., Vita, J.P. y Zamora, J.A. Zaragoza.

Simon, H. (1969). *The sciences of the artificial*. Ed. Mitt-Press. Cambridge, Massachusetts.

Trillas, E. (1995). *Introducción a la lógica borrosa.* Barcelona.

Anexo. Lámina 1.
Situación en el mapa del Antiguo Oriente Próximo de Ugarit.



Mapa realizado por B. Solans y J. Justel.

Lámina 2

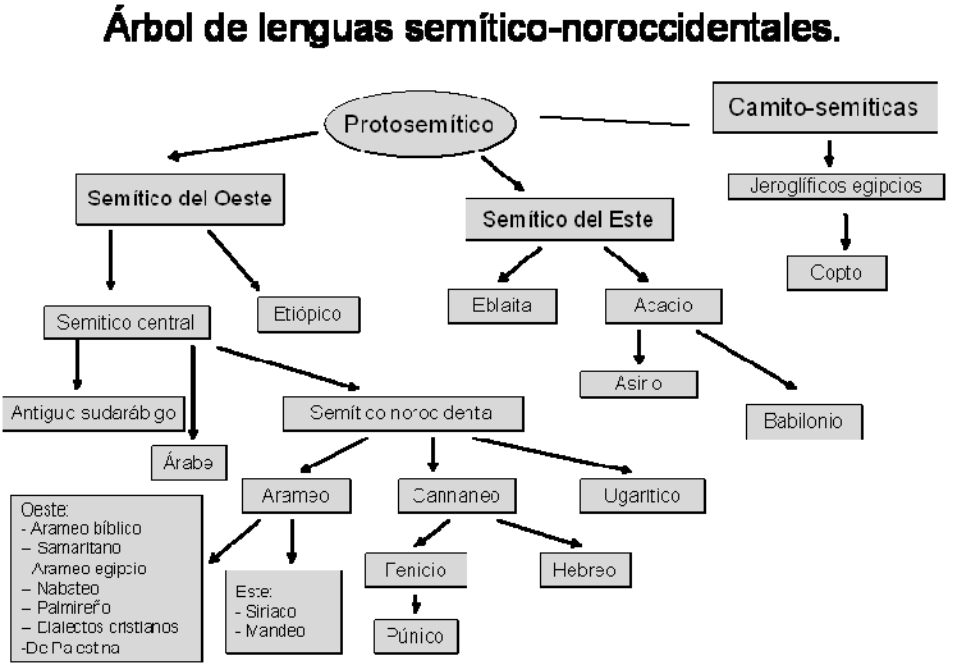
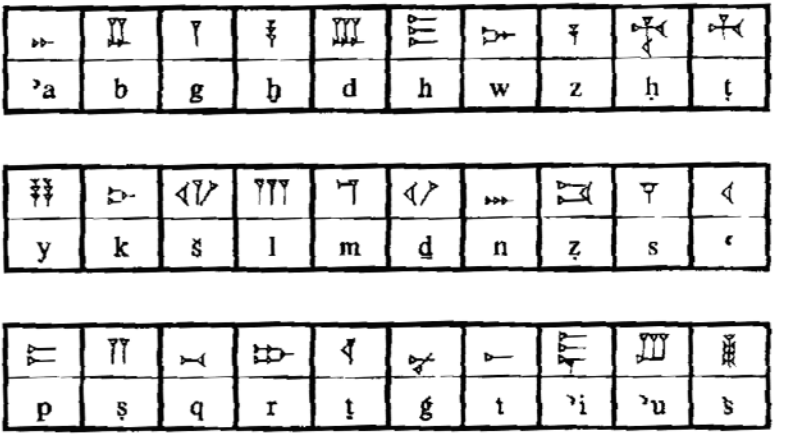


Lámina 3. Alfabeto ugarítico



THE UNIVERSAL SET – A (NEVER FOUGHT) BATTLE BETWEEN PHILOSOPHY AND MATHEMATICS

Reinhard Kahle
Centria and DM,
FCT, Universidade Nova de Lisboa

Früher galt es, als Höhepunkt der Strenge, wenn man sagte:
Ein Begriff ist definiert, wenn man von „jedem" Ding sagen kann,
ob es darunter fällt oder nicht.
David Hilbert, [Hill7, S. 135f]

Abstract. 100 years ago, in 1908, ERNST ZERMELO published his axiomatization of set theory [Zer08]. Although this axiomatization was by no means uncontroversial, it led eventually to the frameworkwhich, nowadays, is generally accepted as *the* foundational framework of mathematics. With ZERMELO'S axiomatization set theory reached its first milestone after its introduction by GEORG CANTOR in the last decades of the 19th century. Cantorian set theory was confronted from its very beginning with certain difficulties, usually classified as *paradoxes*[1]. In the context of these difficulties a discussion started which – ultimately – led to the transition of logic as a philosophical discipline to a subfield of mathematics. While the general discussion of the paradoxes in Cantorian set theory is a well-known story[2], in this note, we will discuss the (potential) function of the *universal set* in this debate, in particular with respect to its consequences for the primacy of philosophy or mathematics over logic.

1. From concepts to sets

We start our consideration with the notion of *concept* which has a dignified history in philosophy since ARISTOTLE. It is one of the fundamental notions[3] of traditional[4] logic[5]. Philosophers discussed a

[1] CANTOR himself didn't classify the contradictory set constructions as paradoxes, but rather as *reduction-ad-absurdum* arguments for the non-existence of these sets, see below.
[2] Cf. e.g., [Gar92].
[3] The other is the notion of *judgment*.

lot, in the context of traditional logic, the difference between the *intention* and *extention* of a concept.[6] Roughly, the intention is given by *all attributes* which the concept inhere; the extention by the *collection of things* which fall under the concept. The extention comes already close to the modern notion of set; however, in traditional logic, concepts are equipped with a *mental criterion*.[7]

It was only BOLZANO who introduced the modern notion of *set*[8] when he explicitely abstracted away from any criterion which relate the elements which each other [Bol76, Vol. IIA, p. 152]:[9]

> The parts, of which a set is composed, define it, and in fact completely and all in the same way.

Although finite and *potentially* infinite sets – in the sense of BOLZANO – were considered in mathematics since ancient times, it was only CANTOR – with BOLZANO'S notion of set at hand – who

---

[4] We use here *traditional logic* as a term for the philosophical discipline of logic as it was established before the rise of modern (mathematical) logic, i.e., Aristotelian logic including its elaboration by the Scholastics. We are not sure to which extent this designation is standardized; the alternative terms *philosophical logic* or *classical logic* would be rather ambiguous, since they are also used within modern logic.

[5] In the following we will restrict *concept* to its technical meaning in traditional logic, as it is expressed, for instance, also by the German word *Begriff* ; the English word *concept* seem to have not only in colloquial language a wider meaning (as the German *Begriff*), but also as (modern) philosophical term.

[6] ARNAULD uses *compréhension de l'idée* and *étendue de l'idée*, respectively, in the *Logique du Port Royal*, [Arn85, I. Partie, Chap. VI], cited according to [Fel02, p. 174]. The standard German terms are *Begriffsinhalt* and *Begriffsumfang*.

[7] In this form, it is explicitely formulated by CASSIRER [Cas10, p. 8]:

> Somit zeigt sich, daß die allgemeine formale Vorschrift [of concept formation] für sich allein nicht genügt, daß vielmehr überall zu ihrer Ergänzung stillschweigend auf eine anderes gedankliches *Kriterium* zurückgegriffen wird.

CASSIRER sees in ARITOTLE' S *metaphysics* the role to provide such a criterion, and assigns the link between logic and metaphysics to the theory of concept.

[8] In fact, BOLZANO did not only introduced the notion, but he also coined the German term "Menge". This term was then taken over by CANTOR (who initially also used "Mannigfaltigkeit" ["manifold"]).

[9] In German (cited in [Fel02, p. 176]):

Die Theile, aus denen eine Menge besteht, bestimmen sie, und zwar vollständig und alle auf einerley Art.

Since CANTOR we speak about *elements* instead of *parts*.

considered proper *transfinite* sets. The discovery[10] of this new mathematical realm marks a turning point in the history of mathematics. But, from the very beginning, the question was raised which sets can be defined consistently. CANTOR'S famous original definition of a set as "any collection M into a whole of definite, distinct objects m (which are called the 'elements' of M) of our perception [Anschauung] or of our thought" turned out to be too vague. It gave rise to several paradoxes, of which RUSSELL' S paradox of "non-self-membership" is the classical example:[11] By use of the basic element relation, we can define the predicate $P(x) := x \notin x$. In any reasonable set theory, the definition of the set $R = \{x | x \notin x\}$ will lead to a contradiction, since $R \in R$ is equivalent to $R \notin R$.

While RUSSELL'S paradox had probably the biggest impact on the discussion of formal systems for logic, it was neither the first nor – in our view – the most significant paradox of CANTOR'S naive set theory (it might be considered as the simplest). In fact, as the most significant one we consider *Cantor's paradox*. One of the first results in Cantorian set theory was what is now called *Cantor s theorem*: The power set $\mathfrak{R}(X)$ of a given set X has higher cardinality as X. However, it has an immediate *paradoxical* consequence:

> There cannot be any *universal set V*, containing all elements of discourse.

Since $\mathfrak{R}(V)$ would have more elements than V, V cannot contain all of them. The "paradox" was discussed by Cantor as early as 1899 in his correspondence with DEDEKIND.[12] But CANTOR didn't consider it as paradox at all. He took it as a *reductio-ad-absurdum* argument for the non-existence of the universal set.

[10] Depending of the reader' s philosophical attitude, one may read "invention".

[11] The paradox was independently discovered by ZERMELO in Göttingen, cf. [RT81]. It was probably known to HILBERT since 1899/1900 (cf. HILBERT' S postcard to FREGE from Nov 7th, 1903: "I believe Dr. Zermelo discovered it three or four years ago […]." [Fre80, p. 51]; German original [Fre76, pp. 79–80].

[12] It was probably known to CANTOR long before, but the correspondence with DEDEKIND is the earliest proof we have; it was published only in his collected works [Can32], edited by ZERMELO, in 1932.

2. Justifying sets

Cantor's rather naive attitude towards set existence was heavily criticized, in particular by David Hilbert, who nevertheless was one of the leading defenders of the new area of set theory. In his address to the International Congress of Mathematicians in 1904, HILBERT said [Hil05a]:[13]

> G. Cantor experienced the mentioned contradiction [of the concept of the totality of all things] and expressed this experience by distinguishing 'consistent' and 'inconsistent' sets. But, inmy opinion, by not establishing a clear criterion for this distinction, I must call his approach in this point one which still leaves latitude for *subjective* judgment and therefore affords no objective certainty.

Still in 1917, HILBERT repeated this critique in more drastic words in a lecture on set theory at the University of Göttingen [Hil17, S. 135]:[14]

> Thus, at one time two mathematical operations shall be admissible, because they do not result in a contradiction [in the generation of the second number class; i.e., real numbers]; but at another time two such processes shall be inadmissible, because a paradox is deduced, and only the success shall decide what is forbidden and what not. […] Such a position is of course indefensible. If we can check a mathematical proof for its admissibility only by the result, then we need no proof at all.

[13] In German (cited in [Can91, S. 436]):
G. Cantor hat den genanntenWiderspruch [den des Begriffs der Gesamtheit aller Dinge] empfunden und diesem Empfinden dadurch Ausdruck verliehen, daß er „konsistente" und „nichtkonsistente" Mengen unterscheidet. Indem er aber meiner Meinung nach für diese Unterscheidung kein scharfes Kriterium aufstellt, muß ich seine Auffassung über diesen Punkt als eine solche bezeichnen, die dem *subjektiven* Ermessen noch Spielraum läßt und daher keine objektive Sicherheit gewährt.

[14] In German:
Das eine Mal sollen also zwei mathematische Operationen erlaubt sein, weil sich keinWiderspruch ergibt [in der Bildung der zweiten Zahlklasse]; das andere Mal aber sollen zwei solche Prozesse unzulässig sein, weil eine Paradoxie folgt, und nur der Erfolg soll entscheiden, was verboten ist und was nicht. [...] Ein solcher Standpunkt ist nat ürlich unhaltbar. Wenn wir einen mathematischen Beweis erst am Resultate auf seine Zulässigkeit prüfen können, so brauchen wir überhaupt keinen Beweis.

HILBERT himself started, in fact, from the traditional *justification* criterion for concepts, and tried to apply it to sets. A first reference we find at the very beginning of his discussion of set theory in the correspondence with CANTOR. In a letter from October 2nd, 1897 CANTOR cites HILBERT with the words:[15]

> The collection [German: "Inbegriff"[16]] of alefs can be taken as a particular well-defined set, since when some thing is given it always must be possible to decide whether this thing is an alef or not; but more is not needed for a well-defined set.

Getting to know the paradoxes, HILBERT realized that this justification is problematic. So, we find in the famous postcard in answer to FREGE'S complimentary copy of his second volume of the *Grundgesetze der Arithmetik* the following passage:[17]

> [Other contradictions] led me to the conviction that the traditional logic is insufficient; in fact, the theory of formation of concepts needs a tightening and refinement; I see as the most essential gap the assumption in the traditional system of logic

[15] Cantor to Hilbert, Oct. 2nd, 1897; in German (cited in [PI87, S. 226] and [Can91, S. 390]):

> Der Inbegriff der Alefs läßt sich als eine bestimmtewohldefinirte Menge auffassen, da doch wenn irgend ein Ding gegeben wird allemal muß entschieden werden können, ob dieses Ding ein Alef sei oder nicht; mehr aber gehört doch nicht zu einer wohldefinierten Menge.

Cantor responded to this notion of "well-defined" set by his distinction of *consistent* and *inconsistent* sets; a distinction which was critized by Hilbert in the citations given above.

[16] The German word "Inbegriff" is obsolete (as logical term) and not easy to translate. FELGNER argues that it was used as "Menge" (or extension of a concept or something like that) but *together with* something in addition, for instance an order relation of the elements. He concludes that the notion of *structure* as introduced by BOURBAKI comes closest to "Inbegriff". [Fel02, p. 174f]

[17] Hilbert to Frege, Nov. 11th, 1903; in German (cited in [Fre76, S. 79f]):

> [AndereWidersprüche] f ührten mich zu der Ueberzeugung, dass die traditionelle Logik unzureichend ist, die Lehre von der Begriffsbildung vielmehr einer Verschärfung und Verfeinerung bedarf, wobei ich als die wesentlichste Lücke im herköm[m]lichen Aufbau der Logik die Annahme ansehe, wonach – das nehmen alle Logiker u. Mathem[atiker] bisher an – ein Begriff bereits da sei, wenn man von jedem Gegenstande angeben könne, ob er unter ihn falle oder nicht. Dies ist wie mir scheint nicht hinreichend.

> that a concept already exists – as all logicians and mathematicians so far agree – when one can decide for every thing whether it falls under the concept or not. As it seems to me, this is not sufficient.

With the same negative attitude, HILBERT repeated this "traditional" definition of set in a lecture on set theory in 1917, [Hil17, p. 135f]:[18]

> In former time, it was considered as maximum of rigor if one says: A concept is defined, if one can say for "every" thing whether it falls under it or not. Now, we have to reject this point of view rigorously.

But, still in 1920, Hilbert uses this definition, saying [Hil20, p. 11f]:[19]

> One considers a set aswell-defined, if for every thing it is determined [later corrected for: "one can decide"] whether it belongs to the set or not.

Although HILBERT considered the traditional definition as "insufficient", due to the irrefutable mathematical results in set theory, he was puzzled with justification for sets. This is best expressed in the following questions, which HILBERT asked in 1905, with Cantor's paradox in mind, [Hil05b, p. 215]:[20]

---

[18] The German original of the first sentence is used as motto at the beginning of this text; the second reads:

> Wir müssen diesen Standpunkt daher jetzt aufs schärfste zurückweisen [...].

[19] In German:

> [M]an betrachte eine Menge als wohldefiniert, wenn von jedem Dinge bestimmt ist [korrigiert statt: "man... entscheiden kann"], ob es zur Menge geh¨ ort oder nicht[.]

Note the nearly literal similarity of this definition with DEDEKIND' S definition in *Was sind und was sollen die Zahlen?* [Ded87, second section]: "[A system (set) *S*] is completely defined, if it is defined for every thing whether it is an element of *S* or not." (In German: "[Ein solches System *S* (oder ein Inbegriff, eine Mannigfaltigkeit, eine Gesamtheit)] ist vollständig bestimmt, wenn von jedem Ding bestimmt ist, ob es Element von *S* ist oder nicht.").

[20] In German:

> Warum ist die Gesammtheit aller Mengen eine unerlaubte, warum die Menge aller reellen Zahlen eine erlaubte Zusammenfassung?

> Why is the totality of all sets not permissible?
> Why is the set of all real numbers a permissible collection?

3. The never fought battle

So far, we did not comment on one particular consequence of the traditional notion of concept:

> The universal concept exists trivially, when concept are characterized by the fact that one can decide whether an element falls under it or not.

Although there are some arguments in traditional logic against such a, essentially *extensional*, characterization of concepts[21], in comparison with the mathematical notion of set, the extensional view is clearly the relevant one. Thus, reducing the *philosophical* position to this extensional view of concepts, we can note three observations (identifying sets with such concepts):

> **1.** The characterization via a decision procedure for the element relation *is* a (philosophical) answer to the question, why a set exists.
> **2.** It works, to some extent, to block paradoxes, first of all, to block RUSSELL'S paradox.[22]
> **3.** However, it trivially justifies the universal set.

In contrast, for the *mathematical* notion of set we have:

To be historically accurate, Hilbert asked these question with respect to *his own paradox*, a paradox which can be considered as a variant of Cantor's paradox; cf. [PK02, Kah06].

[21] As noted in footnote 7, CASSIRER criticized the extensional view as neglecting a mental criterion. He argues particularly against the "universal concept" as a, although definable, but empty concept [Cas10, p. 7]. A similar criticism was done by HUSSERL against FREGE' S explicit identification of concepts with the extension of a concept, cf. [Hus92, p. 122] with reference to [Fre84, footnote on page 80].

[22] It is immediate to see that we have no possibility to decide whether the Russell "set" R belongs to itself or not; ergo, R is not a well-defined set.

1. There is no *a priori* criterion for set existence.[23]
2. Paradoxes can, so far, only be blocked *ex post* by axiomatic restrictions.
3. There is no universal set, due to *Cantor's theorem*.

From this perspective, the decision between the primacy of philosophy or mathematics over logic, between the (traditional) concepts and (modern) sets, can be reduced to the decision about the existence of the universal set. Thus, philosophy and mathematics would have to battle over the universal set. This battle did not take place. Probably because it was not fought on philosophical ground, but on mathematical – and philosophers never entered that ring. In fact, there were philosophers commenting on the problem, as Husserl and Cassirer (cf. footnote 21 above), but none of them succeeded in impede mathematics to take over logic from philosophy. It might be debatable if any of them had actually the intention to do so;[24] but they will have had definitely the standard that entities like concepts or sets have to be justified on philosophical grounds and not only by an axiomatic approach.

So, how set theory won over the traditional concepts on mathematical grounds? Simply by its success. And the success of set theory is, first of all, manifest in the possibility to define the set of real numbers. As a matter of fact, mathematicians would never be willing to give up of the real numbers.[25] With respect to our core issue, the universal set, it is its fate that it is incompatible with the power set axiom, the key tool to reach the reals. And, with the fall of the universal set, the philosophical notion of concept also fell. This

[23] In this context, it is worth noting that ZERMELO'S axiomatization (and any other) may be considered as an answer to the question: "What is a set?" But, it gives a purely normative answer: set is what can be constructed from the axioms. From a philosophical point of view, the axiomatization can hardly be considered as an answer to the question "Why a set exists?"

[24] There were some few open defenders of traditional logic, as, for instance, JACOBY and VON LÖRINGHOFF, cf. e.g., [Jac62], which, as mavericks, fought a completely lost battle.

[25] In the following section we will discuss some alternatives to standard set theory, which in general, however, fail to capture the (desired) mathematical notion of real numbers. And while these alternatives have definitely some attractive features, just because they fail to provide a satisfactory account to the reals, they were abandoned by mathematics.

decision is made on mathematical grounds for the price that question, why a set exists, remains open.[26]

4. Modern theories with universal sets

While Zermelo-Fraenkel set theory ZF is nowadays the generally accepted fundamental theory of mathematics, there are alternative approaches which, in contrast to ZF, allow for a universal set, or something comparable.

In first place, one has to mention VON NEUMANN'S set theory, nowadays called *von Neumann–Bernays–Gödel set theory* NBG (due to the variant axiomatization given by BERNAYS and GÖDEL), cf. [FB58]. In this set theory, sets are distinguished from CLASSES. While every set is a class, there are some classes – like the universal class – which are not sets. For such classes, not all set-theoretic operations are allowed. Very simplified, one can say that classes are allowed to appear on the right side of the element relation, but not on the left side. It is shown that ZF and NBG are essential equivalent in their mathematical expressiveness, such that, for a mathematician, the choice between these two frameworks is rather a matter of taste. Probably due to its "late" introduction, to our knowledge, the classes of NBG were never confronted with the notion of *concept* of traditional logic; and it seems to be rather doubtful that they would serve better than the sets of ZF.

From a philosophical point of view, QUINE'S *New foundation* is an interesting alternative approach to set theory, which allow for a universal set, cf., e.g., [For95]. However, this system is rather neglected from the mathematical side, and one might say that its existence is primarily justified by the (annoying) open question whether it is a consistent framework or not.

The most serious alternative to classical set theory is given by BROUWER'S *intuitionism*. This framework is "anti-set-theoretic" in nature, and, in view of its inventor, it does not require any axiomatization. However, HEYTING was providing one in any case,

[26] This is in striking analogy to the situation in quantum mechanics: Although there is till this day no philosophically satisfying interpretation –the Copenhagen interpretation can hardly be considered as *philosophically* satisfying –, only its success led to the general acceptance of quantum mechanics in physics.

and, starting from this, there are nowadays a couple of variants of intuitionistic and *constructive* theories around. While the existence of a universal set seems to be rarely addressed directly, the notion of the *real numbers* is in the focus of these theories; and, at a rule, they restrict themselves to countable sets of real numbers. As a consequence, the *diagonalization* argument seems sometimes to be put into question; in fact, it is not the diagonalization argument in its core, but second step which is usually added:[27] for every (constructively) given enumerable set *A*, one can *construct*
by Cantor's diagonalization method an element which does not belong to *A*. However, what for some constructivists[28] is not possible, is the "construction" of a *set* containing *all* elements which "in principle" can be constructed by Cantor's diagonalization method. In this view, this method can applied to every enumerable sets, resulting in a new enumerable set, but there is no set as the closure of all of them.[29] Although it is unclear, whether such a constructivism would be happy with a universal set, it seems to escape, at least, from Cantor's paradox.

Finally, we would like to mention FEFERMAN'S framework of *explicit mathematics* [Fef75, Fef79], and the related axiomatic theory of *Frege structures* as modern theories which allow for a universal set (also called *universal type* or *universal class*). These frameworks are of interest, since they allow a certain insight in the interplay of diagonalization and a universal set, at the expense of the power set operation. This is discussed in detail in [Kah0x] (with reference to [Can96]).

[27] I owe the following argument to CHRISTIAN THIEL, Erlangen.

[28] It depending on the particular understanding of constructivism; the argument does not apply to every version, as, for instance, to the presentation of *intuitionism* in [McC05] (where Corollary 2 explicitly states the uncountability of a power set of N.)

[29] In fact, this is the same situation as for Euclid's proof of the "existence of infinitary many primes"– stressed by constructivists over and over again: Euclid only states the version: "Prime numbers are more than any assigned multitude of prime numbers." (Euclid, Elements, Book IX, Proposition 20). The proof – which is in fact only a schematic proof by example – does not speak about any infinite set of prime numbers.

5. Conclusion

In summary, we have the following situation:

The existence of the universal set can be seen as an "experimentum crucis" for the traditional notion of concept. Modern set theory rejects the existence of the universal set by Cantor's theorem. To save it, one would have to give up the power set axiom. Since this axiom is needed for the theory of real numbers – and mathematics would never be willing to give up of this theory – the universal set is banned from the mathematical discourse. As a matter of fact, the philosophical notion of concept became out of fashion, and the mathematical notion of set is generally accepted.

However, the notion of concept came with a philosophical satisfying justification
of its instances; in contrast, set theory is not able to give a philosophical justification for the existence of particular sets. Its justification is just the (mathematical) success.

Thus, since Hilbert's questions

"Why is the totality of all sets not permissible?"

"Why is the set of all real numbers a permissible collection?"

still await an answer, the story of the universal set does not seem to be told to its end.

References

[Arn85] Antoine Arnauld. *La logique ou l' art de penser*. Amsterdam, 1685.

[Bol76] Bernhard Bolzano. *Gesamtausgabe*. Frommann, Stuttgart, edited by J. Berg et al.; Reihe II: Nachlass, Band A7: Einleitung zur Größenlehre und erste Begriffe der allgemeinen Größenlehre. Band A8: Reine Zahlenlehre 1975/76.

[Can32] Georg Cantor. *Gesammelte Abhandlungen mathematischen und philosophischen Inhalts*. Herausgegeben von Ernst Zermelo, Springer, 1932.

[Can91] Georg Cantor. *Briefe*. Herausgegeben von Herbert Meschkowski und Winfried Nilson, Springer, 1991.

[Can96] Andrea Cantini. *Logical Frameworks for Truth and Abstraction*, volume 135 of *Studies in Logic and the Foundations of Mathematics*. North-Holland, 1996.

[Cas10] Ernst Cassierer. *Substanzbegriff und Funktionsbegriff*. Berlin, 1910. Reprint: Wissenschaftliche Buchgesellschaft, Darmstadt, 1994.

[Ded87] Richard Dedekind. *Was sind und was sollen die Zahlen?* Braunschweig, 1887.

[FB58] Abraham Fraenkel and Paul Bernays. *Axiomatic Set Theory*. North-Holland, 1958.

[Fef75] Solomon Feferman. A language and axioms for explicit mathematics. In J. Crossley, editor, *Algebra and Logic*, volume 450 of *Lecture Notes in Mathematics*, pages 87–139. Springer, 1975.

[Fef79] Solomon Feferman. Constructive theories of functions and classes. In M. Boffa, D. van Dalen, and K. McAloon, editors, *Logic Colloquium 78*, pages 159–224. North–Holland, 1979.

[Fel02] Ulrich Felgner. Zur Geschichte des Mengenbegriffs. In Bernd Buldt et al., editors, *Kurt Gödel: Wahrheit & Beweisbarkeit*, volume 2: Kompendium zum Werk, pages 169–185. öbv & hpt, 2002.

[For95] Thomas E. Forster. *Set Theory with a Universal Set*, volume 31 of *Oxford Logic Guides*. Oxford University Press, 1995.

[Fre84] Gottlob Frege. *Grundlagen der Arithmetik*. Wilhelm Koebner, 1884.

[Fre76] Gottlob Frege. *Wissenschaftlicher Briefwechsel*. Herausgegeben von Gottfried Gabriel et al., Meiner, 1976.

[Fre80] Gottlob Frege. *Philosophical and mathematical correspondence*. Basil Blackwell, 1980. edited by G. Gabriel et al.

[Gar92] Alejandro R. Garciadiego. *Bertrand Russell and the Origins of the Set-theoretic 'Paradoxes'* . Birkhäuser, 1992.

[Hil05a] David Hilbert. Über die Grundlagen der Logik und der Arithmetik. In Adolf Krazer, editor, *Verhandlungen des Dritten Internationalen Mathematiker-Kongresses in Heidelberg vom 8. bis 13. August 1904*, pages 174–185. Leipzig, 1905.

[Hil17] David Hilbert. Mengenlehre. Vorlesung Sommersemester 1917, Ausarbeitung (Bibliothek des Mathematischen Instituts der Universit ät Göttingen), 1917.

[Hil20] David Hilbert. Probleme dermathematischenLogik. Vorlesung Sommersemester 1920, Ausarbeitung von N. Schönfinkel und P. Bernays (Bibliothek des Mathematischen Instituts der Universität Göttingen), 1920.

[Hus92] Edmund Husserl. *Philosophie der Arithmetik*, volume 1 of *Gesammelte Schriften*. Felix Meiner Verlag, edited by E. Ströker; reprinted from Husserliana XII edited L. Eley, Martinus Nijhoff, 1970. 1992.

[Jac62] Günther Jacoby. *Die Ansprüche der Logistiker auf die Logik und ihre Geschichtsschreibung*. Stuttgart, 1962.

[Kah06] Reinhard Kahle. David Hilbert über Paradoxien. Pré-publicações, Departamento de Matemática, Universidade de Coimbra, 2006. Preprint Number 06-17.

[Kah0x] Reinhard Kahle. The universal set and diagonalization in Frege structures. 200x. Submitted.

[McC05] D. C. McCarthy. Intuitionism in mathematics. In Steward Shapiro, editor, *The Oxford Handbook of Philosophy of Mathematics and Logic*, pages 356–386. Oxford, 2005.

[PI87] Walter Purkert and Hans Joachim Ilgauds. *Georg Cantor 1845–1918*. Vita Mathematica, 1. Birkhäuser, 1987.

[PK02] Volker Peckhaus and Reinhard Kahle. Hilbert's Paradox. *Historia Mathematica*, 29(2):157–175, 2002.

[RT81] Bernhard Rang and W. Thomas. Zermelo's discovery of the 'Russell paradox'. *Historia Mathematica*, 8:15–22, 1981.

[Zer08] Ernst Zermelo. Untersuchungen über die Grundlagen der Mengenlehre I. *Mathematische Annalen*, 65:261–281, 1908.

ANALOGÍA Y SEMEJANZA EN LA REPRESENTACIÓN E IDENTIFICACIÓN DE LOS CONCEPTOS TEÓRICOS COMO CONCEPTOS COMPLEJOS

Francisco J. Salguero Lamillar
Universidad de Sevilla

El concepto de concepto teórico

No cabe duda de que los conceptos teóricos desempeñan un papel importante en el entramado lógico y cognitivo de las teorías científicas, aunque este papel ha resultado tradicionalmente de difícil encaje en las teorías empiristas del conocimiento desde la antigüedad griega en adelante. Seguramente por ello la discusión acerca de su naturaleza se encuentra en el centro de gran parte de las teorías del significado del siglo XX, principalmente en la filosofía analítica y en todos aquellos sistemas herederos del positivismo lógico.

La noción de *concepto teórico* proviene de la distinción positivista entre aquellos conceptos referenciales cuya denotación depende de la experiencia sensible y aquellos otros conceptos referenciales que no se sustentan en la observación empírica directamente, pero que son necesarios para la consistencia o la capacidad explicativa de una teoría. Se trata, por tanto, de una distinción epistemológica relacionada con la ontología de las teorías científicas en las que intervienen dichos conceptos. Sin embargo, su importancia en la teoría del conocimiento es anterior a los planteamientos positivistas del veinte, pudiéndose rastrear en las teorías del conocimiento atomistas, platónicas, aristotélicas y epicúreas, en los postulados epistemológicos de los gramáticos especulativos medievales o en los planteamientos fundamentales de las filosofías racionalistas y empiristas de la modernidad.

En una teoría dada, los conceptos teóricos se representan mediante los denominados *términos teóricos*, expresiones lingüísticas cuyo significado referencial no es reducible a su representación observacional en la mente del hablante. El positivismo lógico entendió, sin embargo, que los enunciados en los que intervenían

términos teóricos podían ser reducidos a otros enunciados sinónimos en los que todos los términos participantes tuviesen una referencialidad observacional, de modo que la representación mental de su significado se basase exclusivamente en el significado de los términos observacionales correspondientes y las leyes de la lógica. Estas representaciones mentales se denominan *proposiciones protocolares* [Carnap 1928] y su principal cometido es evitar postular una ontología no estrictamente empírica en las teorías científicas.

De este modo, conceptos como "mesa", "cinco" o "claridad" se corresponden con la experiencia que los usuarios de estos términos han tenido de tableros que se sustentan sobre patas y se elevan del suelo para servir de apoyo, grupos de objetos en cantidad igual a los dedos de una mano o la manifestación visual de la luz del sol mientras éste se encuentra presente en el firmamento. A partir de estas experiencias, extrapolamos otros perceptos para construir conceptos basados empíricamente como "sólido" (como la mesa, frente a "líquido" como el agua), "cien" (en cantidad igual a los dedos de veinte manos) o "calor" (sensación táctil que acompaña a la sensación visual provocada por la manifestación del sol cuando se encuentra presente en el firmamento), y podemos interpretar enunciados como "el agua está hirviendo" traduciéndolos a sus correspondientes enunciados protocolares: "hay un objeto tal que ese objeto es líquido y al que llamamos agua y que el calor ha hecho que alcance una temperatura en cantidad igual o superior a 100 en la escala correspondiente". Cada uno de estos términos requiere la comprensión de otros conceptos más "primitivos" o derivados de los anteriores como el concepto de "cantidad", el concepto de "veinte", el de "temperatura" o el concepto de "firmamento". Los más primitivos como "cantidad" pueden vincularse a percepciones, en tanto que los derivados dan lugar a conceptos cada vez más complejos, necesarios para comprender lo que percibimos. Algunos de estos últimos conceptos complejos serían los denominados conceptos teóricos.

El problema se plantea cuando surgen términos y enunciados que no son reducibles a su correspondiente forma protocolar. En estos casos el positivismo optaba por considerar tales expresiones como carentes de sentido (*nonsenses*) y por tanto inapropiadas para el discurso científico. Sin embargo, muchos de estos sinsentidos admiten una interpretación clara en el lenguaje ordinario, en tanto que los enunciados protocolares son considerados a veces por los hablantes

como expresiones incorrectas o poco afortunadas desde el punto de vista de la gramática de una lengua. Esto llevó a algunos teóricos del positivismo lógico como L. Wittgenstein a replantearse drásticamente la naturaleza del significado y la representación cognitiva de los conceptos en la mente de los hablantes [Wittgenstein 1953].

Si partimos de la idea de que los conceptos son representaciones mentales asociadas a diversos estados de la mente – entre los cuales se hallan también, sin lugar a dudas, las palabras que los nombran – entonces deberemos considerar cómo se originan y cómo se relacionan entre sí estos conceptos. En este sentido habremos de otorgar un importante papel al *léxico mental* en su definición, pues son la vía más directa para poder acceder a ellos, por lo que podremos analizarlos como estados de información en la mente del individuo que remiten a campos cognitivos más o menos amplios y relacionados entre sí. Estas relaciones deberían poderse modelizar en una semántica de mundos posibles que permitiera interpretar situacionalmente el contenido cognitivo de tales conceptos, por ejemplo.

Abordamos, por tanto, varios problemas aquí. El problema de la definición representacional de los conceptos, el problema de la relación de estas representaciones con el léxico mental y el problema de los modelos de interpretación del léxico mental, quedando éste último para un trabajo de otra índole.

Conceptos como representaciones

La definición representacional de los conceptos nos lleva a plantearnos cuáles son los contenidos mentales asociados a la identificación de un concepto y cómo es su representación en términos cognitivos. A veces, podemos creer que se trata de imágenes, de "fotografías psíquicas" del referente real del concepto. Esto está bien si estamos hablando de conceptos como "La Tour Eiffel" o "la Puerta del Sol en Nochevieja", siempre que retengamos en la retina una imagen de ambas. Pero ¿qué ocurre en la mente cuando se reconoce un estilo musical o pictórico, por ejemplo? Cuando estamos escuchando una pieza musical y decimos: "es un *concerto grosso* barroco" o bien "es una sinfonía romántica"; o quizás estamos mirando un cuadro y nos parece "impresionismo francés" o

"hiperrealismo fantástico", ¿qué es lo que relacionamos en nuestra mente con estos conceptos?

Es posible considerar la posibilidad de que la relación se establezca entre una percepción directa y el significado de una cierta expresión lingüística. Es decir, la representación cognitiva del concepto "sinfonía romántica" sería la correspondiente al significado del SN, un nombre modificado por un adjetivo que restringe el conjunto de entidades referidas por el propio nombre. De este modo, todos los conceptos estarían ligados a una representación referencial léxica que dependería del significado léxico y de la gramática de la lengua en la que se expresa ese concepto. Sin embargo, es evidente que esto sólo desplaza el problema desde el nivel epistemológico al nivel semiótico o del significado, por lo que estaríamos ante una *petitio principii*, pues el significado estaría siendo entendido como la representación de la referencia de una expresión lingüística, en tanto que lo que queremos saber es precisamente cómo se representan en la mente del hablante estas referencias y cómo se relacionan entre sí.

No cabe duda de que el léxico es una de las formas en que la mente humana retiene, codifica y expresa el conocimiento adquirido sobre realidades extralingüísticas y, a veces también, extramentales, que no dependen en última instancia de cómo se estructure la lengua. Sin embargo, la cuestión acerca de cómo relaciona la mente conceptos y etiquetas es básica para poder establecer modelos de interpretación del léxico mental y su relación con los referentes que denota, por lo que es preciso representar el conocimiento lingüístico que subyace al significado léxico. De otro modo, la mera teoría referencial del significado no nos llevaría más lejos de lo que nos lleva la definición léxica de las palabras en el diccionario, lo que no bastaría para conseguir modelos útiles de representación del discurso.

Los procesos por los cuales un individuo construye su léxico mental no pueden variar demasiado con respecto a otros individuos de su grupo social, so pena de que cada hablante acabe desarrollando un lenguaje privado incomprensible para los demás miembros del grupo por sus implicaciones semánticas únicas dentro del sistema de la lengua. Los mecanismos que dan lugar al léxico mental, por tanto, deben ser compartidos por todos los miembros de una comunidad de hablantes, lo que supone que deben ser mecanismos básicos de comprensión y fijación de los conceptos los que intervienen en estos procesos, sin que ello suponga en ningún caso que pensemos que estos

mecanismos sean simples o que el mapa mental de todos los hablantes de una lengua tenga que coincidir plenamente.

Un estudio del léxico conocido por un grupo de hablantes y la rapidez con la que estos lo relacionan con unos conceptos o unos significados concretos nos da una pista importante acerca del funcionamiento del léxico mental y de las categorías conceptuales correspondientes. Así, por ejemplo, parece que los seres humanos tendemos a considerar ciertos significados de una palabra más básicos que otros y a categorizar las palabras mismas según una jerarquía conceptual bien definida en la que pueden identificarse prototipos semánticos compartidos por una comunidad de habla [Rosch 1983]. En el caso de las palabras que designan color, por ejemplo, los hablantes tienden a conceptualizarlas según un patrón visual muy similar, de forma que el color "rojo" denota mentalmente para la mayoría de los hablantes de español una mancha regular de un tono brillante de ese color (más cercano al carmesí que al burdeos y muy alejado de los tonos rosáceos, por ejemplo), aunque luego estemos dispuestos a usar ese término genérico para referirnos a otras tonalidades de rojo.

Así, las palabras con carga semántica denotacional se supraordenan en grupos encabezados por un hiperónimo que hace las veces de etiqueta conceptual. Pero los hipónimos dependientes de estas etiquetas no se encuentran al mismo nivel de dependencia entre sí y con respecto al hiperónimo, sino que algunos son "mejores" ejemplos del significado conceptual general que otros. Por ejemplo, para los hablantes de inglés, "robin" (petirrojo) se acerca más al significado conceptual de "bird" (pájaro, ave) que "crane" (grulla), y éste muchísimo más que "hen" (gallina) o "penguin" (pingüino). De esta forma, en el léxico mental de un hablante tipo, petirrojo, gorrión o paloma se acercarían más al concepto de pájaro que pavo o gallina. Lo mismo ocurre con los conceptos de "verdura" "ropa" o "mueble", para los que también existiría un prototipo en el léxico mental.

Rosch concluye que las palabras se relacionan en la mente del hablante por su significado, de modo que nunca se encuentran aisladas en nuestra memoria, sino que forman parte de redes léxicas con distintas conexiones sistemáticas entre ellas. De aquí podemos suponer que el acceso al significado léxico y su recuperación se produce en relación con la etiqueta genérica, con el hiperónimo que usamos para referirnos al concepto más general que otorga significado

a los términos sobre los que ejerce su "parecido de familia", siguiendo la propuesta de Wittgenstein para explicar el significado de conceptos complejos.

Esta noción de "parecidos de familia" (*Familienähnlichkeiten*) [Wittgenstein 1953:§67] no sólo nos ayuda a comprender el comportamiento de las redes léxicas mentales, sino que puede acercarnos a los mecanismos que impulsan la creación e identificación de conceptos en relación con el léxico mental por su vinculación con el razonamiento analógico.

La analogía como fuente del léxico mental

Si exploramos la idea de los enunciados protocolares del positivismo lógico, nos encontramos con que los términos teóricos se entienden como etiquetas que usamos para referirnos a metáforas relacionadas en su origen con nuestra percepción directa de la realidad. El concepto de "átomo" no sería más que una metáfora de nuestra percepción de la divisibilidad de la materia, como el concepto de "Dios" sería una metáfora de nuestra experiencia de la generación de vida y las causas y los efectos a ella asociados. Claro que esto mismo podría decirse de cualquier concepto supraordenado que funcione como hiperónimo léxico en nuestro léxico mental con respecto a otros conceptos dependientes entre los que podamos encontrar algún tipo de parecido de familia.

La analogía, por tanto, estaría en la base de la formación de conceptos teóricos porque es el motor de los procesos metafóricos o metonímicos que les otorgan sentido en la red léxica mental en la que se encuentran. Esta tesis puede verse refrendada por el análisis del significado de aquellos términos o unidades léxicas cuyo significado no puede considerarse estrictamente denotacional, como los conceptos abstractos o algunas unidades fraseológicas. Para entender el significado de estas expresiones, bien podemos postular, siguiendo a [Fauconnier 1997], la existencia de ciertas funciones parciales que relacionan dominios.[1] Estas funciones parciales (*partial mappings*)

[1] Una función es una relación entre elementos de un conjunto (dominio fuente) y los de otro conjunto (dominio destino). Cuando a todos y cada uno de los elementos del dominio fuente les corresponde algún elemento del dominio destino, decimos que la

permiten el desarrollo de procedimientos y principios, de aplicación en un amplio abanico de fenómenos semánticos y de razonamiento, siendo el discurso la vía para investigar estos fenómenos que se analizan lingüísticamente como parte de la semántica, de la pragmática o de la teoría de la argumentación:

> "[…] *los mismos principios y operaciones de 'mapping' funcionan en la semántica elemental, en la pragmática y en el llamado razonamiento de alto nivel. El análisis del tiempo, de la referencia, de la presuposición y los contrafácticos está íntimamente ligado al de los 'mappings' analógicos, las conexiones conceptuales y la construcción del discurso, que por su parte es inseparable del entendimiento de la metáfora y la metonimia, la estructura narrativa, los actos de habla, la retórica y el razonamiento general*". [Fauconnier 1997:5]

Incluso la función referencial del lenguaje que se da cuando las expresiones lingüísticas reflejan sucesos y situaciones objetivos está sujeta a elaboradas construcciones cognitivas, de modo que ninguna expresión se relaciona directamente con su referencia, sino siempre a través de estas estructuras. Dichas construcciones pueden describirse mediante funciones de asignación (*mappings*), que no sólo se refieren a fenómenos como la metáfora o la analogía, sino que se encuentran en el centro mismo de la semántica natural y del razonamiento cotidiano. Según [Fauconnier 1997:9ss], estas funciones pueden ser de diferentes tipos.

> **1. Funciones de proyección (*projection mappings*).** Como en el caso de las metáforas, en las que se proyecta parte de la estructura de un dominio fuente (*source domain*) en otro dominio destino (*target domain*). Por ejemplo, la proyección de las estructuras espaciales en las temporales, en la metáfora EL TIEMPO ES COMO UN VECTOR ESPACIAL. De igual modo, [Lakoff 1987] muestra cómo inferencias construidas en un dominio sensible pueden transferirse a un dominio abstracto como el de la lógica, combinándose las funciones de proyección para dar lugar a nuevos significados. Éste es el caso de las

función es total. Si a alguno de los elementos del dominio fuente no le corresponde ningún elemento del dominio destino, entonces la función es parcial.

metáforas VER ES COMO TOCAR, y SABER ES COMO VER, que se combinan con uno de los sentidos de la preposición "*over*" del inglés para explicar el significado del verbo "*to overlook*": la línea de la visión pasa sobre el objeto (por encima de él, sin tocarlo) y, al no haber contacto, no es visto y, por tanto no es conocido, por lo que no es notado, es pasado por alto (*overlooked*).[2] En cambio, el verbo "*to look over*" tiene un significado bien diferente. [Sweetser 1990] ha establecido también una relación de este tipo con respecto a verbos modales como *may*, *must*, *can*, que pueden expresar leyes de la naturaleza, normas sociales (deónticas) o la necesidad y posibilidad lógicas.

2. Función pragmática (*pragmatic function mapping*). La metonimia y la sinécdoque son "*mappings*" de este tipo. Estas funciones relacionan objetos de una categoría con los de otra categoría diferente, identificándolos referencialmente: PINTORES-CUADROS, AUTORES-LIBROS, PACIENTES-ENFERMEDADES (v. gr.: "Este Picasso me ha costado una fortuna", "Estoy leyendo al mejor Onetti", "La úlcera de la habitación 315 quiere café").

3. Funciones de esquema (*schema mappings*). Se trata de esquemas generales, marcos o modelos que se usan para estructurar una situación en un contexto, como hace ver [Langacker 1987, 1991] con respecto a las construcciones gramaticales y el vocabulario utilizado. Valgan los siguientes textos como ejemplo de cómo un mismo esquema conceptual puede aplicarse a dos funciones de proyección diferentes:

- Mapping basado en la metáfora LA SOCIEDAD ES COMO UN EDIFICIO (esquema conceptual básico EDIFICIO como prototipo de ESTRUCTURA).

"Las **estructuras** sociales se **basan** en **cimientos** sólidos que las hacen duraderas de generación en generación. La deconstrucción postmoderna consiste solamente en derribar **muros, tapiar puertas** y abrir ventanas, de modo que cambian

[2] En este sentido, la frase del español "pasar por alto"–en el sentido de no darse cuenta, no percibir o no notar–puede explicarse perfectamente por estas mismas funciones de proyección que dan lugar a las metáforas mencionadas.

> el **espacio** y las relaciones, pero no los **fundamentos**, que siguen siendo los mismos desde que existen grupos humanos complejos".
>
> - Mapping basado en la metáfora LA MÚSICA ES COMO UN EDIFICIO (esquema conceptual básico EDIFICIO como prototipo de ESTRUCTURA).
>
> "Con su música, John Cage **abre una puerta a otros ámbitos** de la creatividad al margen de la textura del sonido, la melodía o el ritmo. Sus composiciones se **basan** en la alternancia de **espacios llenos y vacíos**, de sonidos y silencios que van **construyendo un ambiente** en el que habitan sensaciones e imágenes que **ocupan** cada **centímetro cúbico**, cada uno de sus **rincones**".

Según Fauconnier, las funciones parciales así definidas (*mappings*) sirven para construir y enlazar lo que él denomina *espacios mentales*, estructuras parciales que usamos en el discurso para dividirlo en estructuras de conocimiento. De esta forma, se hace corresponder un marco cualquiera como el marco COMPRAR y VENDER con un espacio mental, de modo que cada uno de los elementos del marco se relacionan con objetos del espacio mental mediante algunas de las funciones mencionadas antes. Estas funciones son centrales en cualquier construcción cognitiva o interpretación semántica o pragmática del lenguaje. Por ejemplo:

> **1. Contrafácticos.** Los contrafácticos son oraciones del tipo "Si X fuera Y, entonces Z". Así, por ejemplo, las oraciones "Si las piedras hablaran, contarían historias muy interesantes" o "Si yo fuera su padre, no le permitiría hacer eso" son enunciados contrafácticos. En ellos, la relación de "*mapping*" se establece entre, al menos, dos situaciones: una situación real (las piedras no hablan, yo no soy su padre) y otra imaginaria, que pueden ser analizadas desde un punto de vista semántico como mundos posibles alternativos.
>
> **2. Analogías, metáforas y sistemas conceptuales.** En una metáfora como UN ORDENADOR ES COMO UN ORGANISMO, se establece una relación entre los dominios biomédico y de la computación, de forma que el término "virus",

por ejemplo, adquiere un significado nuevo al ser aplicado a los programas que dañan los sistemas computacionales. La analogía entre un virus como microorganismo y un virus como programa informático desempeña un papel importante en la construcción de nuevos significados y la estructuración de nuevos campos semánticos. Este papel de las metáforas y las analogías es fundamental desde un punto de vista sincrónico y también diacrónico para explicar el origen de ciertos significados y usos de expresiones léxicas o unidades fraseológicas, así como la estructuración de nuestras redes conceptuales relacionadas con dichas unidades lingüísticas. Un sistema conceptual de este tipo puede evolucionar atendiendo a funciones de proyección metafóricas y analógicas siguiendo estas etapas [Fauconnier 1997:19-25]:

> **Analogía e inducción del esquema.** Una analogía como la anterior permite al hablante inducir un esquema abstracto y general en el que interpretar otras expresiones relacionadas con la metáfora UN ORDENADOR ES COMO UN ORGANISMO, expresiones como "disco duro infectado", "epidemia informática", "vacunar el ordenador", etcétera. En realidad, este tipo de esquemas generales se desarrolla también para comprender cómo afecta un virus "de verdad" a un organismo, pues la mayoría de los hablantes carece de conocimientos de biología y medicina suficientes, lo que no les impide entender al médico que les explica que sufren una infección y sus consecuencias.
>
> **Categorización y nueva estructura conceptual.** En una segunda etapa, la analogía permite categorizar el dominio destino, clasificando sus componentes. Siguiendo el ejemplo anterior, permite clasificar los programas como virus, vacunas, desinfectantes, inocuos, etc. Así, el vocabulario del dominio fuente pasa a utilizarse en el dominio destino de forma que se pierde la noción de la analogía inicial, estableciéndose un uso metafórico en el que los virus informáticos ya no son "como si fueran virus", sino que quedan conceptualizados con las mismas propiedades infecciosas o epidémicas que los virus reales.

Otorgar nombres y proyectar una estructura. Esta función de proyección (este "*mapping*") permite la transferencia del vocabulario asociado desde el dominio fuente al dominio destino, nombrándolo y estructurándolo de forma análoga. Por tanto, no es que la palabra "virus" haya adquirido un nuevo significado metafórico y que el hablante se olvide de esta relación original una vez que la metáfora esté "muerta" y se produzca la polisemia, sino que la relación de "*mapping*" siempre permanece viva y funcionando gramaticalmente, aunque el hablante no sea consciente de ello. De hecho, cuando el hablante se refiere a virus informáticos, epidemias, vacunas o desinfecciones del ordenador ya no tiene en mente el dominio fuente bio-médico. Lo que funciona aquí es el esquema general inducido a partir de la función de proyección establecida entre ambos dominios. Por tanto, el vocabulario puede adquirir rasgos en el nuevo dominio que están ausentes en el dominio fuente, pudiendo entonces ocurrir varias cosas:

***Blending* e integración conceptual.** *Blending* es una operación cognitiva [Fauconnier & Turner 1994] que consiste en integrar estructuras parciales de dos dominios separados en una única estructura con nuevas propiedades dentro de un tercer dominio. Por ejemplo, se crearía en la mente del hablante una categoría de "virus" de la que los virus biológicos y computacionales serían subcategorías. En este caso, el nuevo dominio podría aplicarse sobre otros dominios destino dando lugar a nuevas relaciones de "*mapping*" (v. gr.: virus sociales o mentales que destruyen ideas, principios morales, que se propagan y mutan, etc.).

Polisemia motivada. Puede ocurrir, sin embargo, que los dos dominios permanezcan diferenciados, de forma que el vocabulario compartido adquiera más de un significado sin que el hablante sea consciente de la relación analógica o metafórica original.

Divergencia y extinción. Se da cuando permanece el vocabulario pero desaparece el vínculo conceptual entre

los términos, o también cuando el dominio fuente cambia su vocabulario, mientras que éste permanece en el dominio destino. Éste es un fenómeno muy habitual en el cambio lingüístico.

3. Unidades fraseológicas y expresiones comunes que dependen de *mappings* complejos. Algunos idiomatismos y unidades fraseológicas requieren de funciones complejas para su correcta interpretación al margen de la literalidad. Por ejemplo, considérese la expresión inglesa "*to catch up*" en la oración

> *I can´t catch up with myself* (literalmente "no puedo alcanzarme a mí mismo").

Este ejemplo sólo es interpretable sobre un "*mapping*" metafórico que permite ciertas construcciones cognitivas: EL TIEMPO ES COMO EL ESPACIO, una función que asocia conceptos temporales y espaciales (los momentos se asocian a puntos en el espacio y los eventos se asocian a momentos). Esto da lugar a interesantes idiomatismos como "estar (o ir) con los tiempos" ("*to keep up with the times*"), "adelantarse al propio tiempo" ("*to be ahead of one´s time*"), etc. Estas construcciones cognitivas no son nada triviales, pues en ellas se proyectan al menos dos niveles de movimiento en la línea espacial del tiempo: se mueve el propio tiempo y se mueven los individuos en un plano paralelo al del tiempo. Todas estas funciones de proyección permiten interpretar el enunciado (1) con el significado de "no consigo ponerme en el nivel de ejecución de tareas que yo mismo me había marcado en mi agenda". Es como si el individuo al que se refieren los pronombres "I" y "myself" se hubiese escindido en dos: una entidad que "viaja" en la línea de movimiento del tiempo y otra que viaja en la línea de movimiento de los eventos y las acciones que se van realizando. En el momento en que ambas líneas confluyan, cobrará sentido la expresión "alcanzarse a uno mismo" ("*to catch up with oneself*").

Podemos concluir entonces que las estructuras necesarias para construir e interpretar conceptos complejos se sustentan en la analogía

como la forma más simple de otorgar sentido al léxico. Estas estructuras de *mapping* nos permiten explicar de un modo razonablemente sencillo las relaciones de dependencia que el léxico mental mantiene con los conceptos generales de los que depende y por qué podemos interpretar una nueva expresión nunca antes oída o una unidad fraseológica cuyo significado no se resuelve composicionalmente. Del mismo modo nos permite ubicar los conceptos teóricos como conceptos complejos en la red léxica correspondiente, que es donde cobran sentido.

Conceptos y situaciones

El modo en que vamos adquiriendo nuestros conceptos e implementándolos cognitiva, lógica o lingüísticamente tiene mucho que ver, desde nuestro punto de vista, con la naturaleza del conocimiento humano. Esencialmente, todos los conceptos son un haz de analogías y situaciones en las que nuestra experiencia se ha encontrado con la necesidad de categorizar entidades, percepciones, sensaciones, etc. Pensar se convierte muchas veces en pasar de un concepto a otro usando estas analogías o las situaciones características que nos permiten reconocer un conjunto de perceptos como parte de una categoría más general o más amplia.

La experiencia nos permite agrupar nuestras percepciones en estructuras de un nivel cada vez más alto. Esto implica que un mayor número de categorías generales hacen de nuestro pensamiento un proceso más homogéneo a medida que nuestras experiencias van siendo mayores. A la vez, las categorías de un nivel superior nos permiten percibir fenómenos que de otro modo quedan ocultos o no son observados como relevantes desde un punto de vista cognitivo o epistemológico, por lo que ampliamos la ontología necesaria para interpretar las nuevas experiencias que podamos ir teniendo.

Estas categorías de nivel superior acaban por dar lugar a categorías abstractas del pensamiento que son necesarias para poder interpretar y agrupar cada vez más hechos y observaciones por analogía con aquellas otras observaciones a partir de las cuales se han constituido estructuralmente. Así, por ejemplo, el proceso perceptivo visual puede concebirse como un ejemplo del modo en que las categorías mentales se disparan o nos vienen las palabras a la mente

cuando observamos una escena o una situación. Pero la representación mental de las situaciones requiere la elaboración de conceptos complejos en la mayor parte de los casos. Estos conceptos complejos admiten etiquetas léxicas o fraseológicas, de modo que el *léxico mental* se elabora en la mente del hablante a partir de los conceptos que usa para representarse mentalmente situaciones de una cierta complejidad.

De este modo, nuestra noción de referencia se modifica sustancialmente, pues las palabras pasan de describir objetos a referirse a situaciones que pueden definirse como descripciones parciales del mundo que satisfacen un concepto complejo. La mente humana siempre está predispuesta a categorizar estas situaciones mediante procesos analógicos basados en los datos empíricos aportados por la observación, agrupando aquellas que poseen ciertas similitudes mediante los mismos conceptos verbalizables en palabras, frases o enunciados. De aquí que el lenguaje se convierta en el mecanismo principal para la adquisición y la estructuración de las categorías complejas en la mente.

Sin embargo, las categorías complejas de la mente son entidades de límites borrosos. No existe una correspondencia única y absoluta entre nuestras observaciones y una categoría ya adquirida, sino que es preciso ir adecuando los perceptos a las categorías que se posee mediante procesos de analogía. Estos procesos permiten la elaboración de *mappings* mentales mediante los cuales se relacionan las experiencias, observaciones, sensaciones con las categorías complejas necesarias para su clasificación e interpretación. Dichas relaciones de *mapping* se etiquetan mediante el léxico mental, por lo que pueden ser recuperadas fácilmente cada vez que se las requiere para interpretar una nueva situación o un nuevo conjunto de experiencias u observaciones. Cada uno de los ítems del léxico mental de un hablante tiene como referencia aquellos objetos o realidades nombrados mediante la misma palabra o frase, independientemente de la naturaleza de los mismos. Y el *mapping* desarrollado en la mente del hablante garantiza que la estructura mental subyacente a todos ellos es la misma para un mismo individuo en diferentes momentos o, incluso, para un mismo grupo de hablantes.

Una reinterpretación de la *hipótesis Sapir-Whorf* y las diferencias léxicas que podemos encontrar entre las distintas lenguas en campos semánticos próximos pueden ilustrar perfectamente esta

argumentación. Si bien es cierto que en su sentido más fuerte la *hipótesis Sapir-Whorf* vincula en exceso léxico, percepción y cultura, no por ello es completamente rechazable el principio básico sobre el que se sustenta: que las etiquetas léxicas de una lengua siempre remiten a referentes culturales observables por un grupo de hablantes. Es claro que en el día a día del uso del lenguaje, los hablantes etiquetan los objetos de su experiencia usando el léxico disponible en su lengua. Estas etiquetas se usan para objetos y experiencias únicos, no necesariamente observados con anterioridad (el autobús en que nos subimos puede estar haciendo su primer recorrido después de salir de fábrica, el coche que pasa en ese momento puede no haber sido percibido por nosotros nunca antes, el árbol que observamos por la ventanilla no lo habíamos visto nunca porque ese día hacemos un recorrido nuevo, etc.). Pero las etiquetas que les otorgamos a estos objetos ("autobús", "coche", "árbol") descansan menos en definiciones abstractas de las categorías correspondientes que en la comparación que hacemos con los mejores representantes de cada una de las categorías observados con anterioridad. Si nos falta la etiqueta eso no significa que desconozcamos la categoría, pero sí que observaremos los objetos correspondientes en función de otras comparaciones o analogías que nos permitan clasificarlos en relación con nuestra experiencia anterior. Así, si la tribu Dani de Nueva Guinea no posee en su vocabulario palabras que denoten color salvo por "blanco" y "negro", eso no significa que no perciban el rojo o el verde o que no puedan concebir categorialmente la diferencia entre objetos de estos colores y los objetos blancos o negros, pues son capaces de clasificarlos atendiendo a esta cualidad. Lo que ocurre es que el *mapping* conceptual que da lugar al léxico de esta lengua distribuye los conceptos de color de forma distinta de como quedan distribuidos en español, por ejemplo.

Por la misma razón, que en español no tengamos una palabra que signifique "llevar un objeto cogido por su extremidad", pero en la lengua maya tzeltal sí (*lip'*), o que no haya una etiqueta léxica simple en nuestro léxico mental que distinga "andar con un ruido de chapoteo a través de un sitio embarrado" de "andar con un ruido de ramas rompiéndose", mientras que los bantúes shona distinguen claramente entre "*chakwair*" y "*chwakatik*", no significa que los hablantes de

español no podamos comprender estos conceptos complejos ni adaptarlos a nuestro léxico mental.[3]

Parece claro, por todo ello, que una categoría compleja puede ser representada de formas diversas en el lenguaje. El léxico es la fórmula más directa de representación de dichas categorías, distinguiéndose en él al menos dos facetas diferentes:

- Una faceta semántica, en la que se codifican aspectos ontológicos de las categorías representadas.
- Una faceta gramatical, en la que se codifican aspectos relativos al tipo de categoría (si es una entidad, un conjunto de entidades, una relación, una situación, etc.).

Pero además del léxico, las lenguas tienen otros recursos de representación de las categorías complejas que generamos en nuestra mente. Éstas pueden representarse mediante unidades fraseológicas compuestas por más de una palabra, pero que se ubican en el léxico mental en niveles similares a los de las palabras en relación con la etiqueta general de la que dependen los demás items léxicos; o bien pueden ser representadas mediante enunciados completos, cuya referencia es una situación o la descripción de un evento.

En cualquiera de estos casos, las categorías complejas quedarían etiquetadas lingüisticamente, lo que nos permite acceder a ellas rápidamente cuando las necesitamos mediante procesos similares a los que aplicamos a las redes léxicas de nuestro léxico mental. La analogía se convierte en el mecanismo fundamental del proceso de elaboración de *mappings* mentales, lo que nos permite ir desde las palabras más simples a las palabras compuestas, las frases o los sintagmas sin ser conscientes en ningún momento de los límites reales de nuestro léxico mental.

[3] Por tanto, la hipótesis Sapir-Whorf acierta al vincular cultura, léxico y significado, pero no acierta cuando adelanta que un hablante no puede comprender un concepto complejo para el que no exista una etiqueta léxica en su lengua. De todas formas, los ejemplos anteriores son ejemplos radicales, porque puede comprobarse cómo diferentes culturas tienden a categorizar objetos o situaciones similares de la misma manera, por lo que la representación mental que nos hacemos de esos objetos y de esas situaciones trasciende las diferencias sociales o culturales existentes entre los hablantes que se los representan.

Los items léxicos son, pues, etiquetas para el núcleo común de aquellos objetos o realidades nombrados mediante la misma palabra o frase. Sea lo que sea lo que esos objetos o realidades comparten, la estructura mental que hay detrás de un término apunta a ello [Hofstadter 2001:535]. El dominio de los conceptos complejos referidos por el lenguaje podría modelarse entonces como un espacio semántico de límites borrosos, cuya indefinición se debe a las sutilezas de las situaciones de *mapping* sobre otras situaciones sobrevenidas y la facilidad humana para establecer analogías entre ellas. Incluso los conceptos teóricos basan su significado en las percepciones que disparan las analogías necesarias para ubicarlos en el espacio semántico correspondiente. Estas percepciones no son azarosas ni neutras desde el punto de vista de la intencionalidad, sino que se encuentran dirigidas hacia la adecuación a conceptos de alto nivel que ya poseemos y con los que nuestras percepciones o creencias se acomodan bien. En el caso de que estos conceptos no encajen con nuestras percepciones o con las hipótesis semánticas que realizamos para la interpretación de una categoría o de un fragmento de discurso, entonces volvemos atrás y buscamos una nueva forma de percibir o de encajar dicha categoría en un espacio mental. Los atractores perceptivos que se encargan de ubicar nuestras percepciones, conceptulizándolas en un determinado espacio mental, son fundamentalmente las experiencias privadas (recuerdos personales), las experiencias públicas compartidas obtenidas de los medios de comunicación y de la educación y la cultura, y el propio léxico estándar de la lengua de un individuo [Kanerva 1988]. Esto significa que nuestro lenguaje y nuestra cultura nos proveen de las etiquetas necesarias para realizar las funciones de atractores perceptivos sin los cuales no podríamos filtrar ni clasificar la realidad en conceptos complejos de alto nivel, lo que convertiría a ciertos acontecimientos y fenómenos de nuestro entorno en “invisibles”; y al contrario, aquellos conceptos que encajaran con nuestros atractores perceptivos se iluminarían semánticamente, adquiriendo su significado de los mapas conceptuales que la analogía permite construir sobre nuestro léxico mental.

Referencias

Carnap, R. (1928). *Der Logische Aufbau der Welt*. Leipzig: Felix Meiner Verlag.

Fauconnier, G. (1997). *Mappings in Thought and Language*. Cambridge: Cambridge University Press.

Fauconnier, G. & Turner, M. (1994). "Conceptual projection and middle spaces". *UCSD Cognitive Science Technical Report*.

Hofstadter, D. (2001). "Analogy as the core of cognition" en D. Gentner, K. J. Holyoak, & B. N. Kokinov (eds.) *The Analogical Mind: Perspectives from Cognitive Science*, Cambridge MA: The MIT Press/Bradford Book, pp. 499-538.

Kanerva, P. (1988). *Sparse distributed memory*. Cambridge, MA: MIT Press.

Langacker, R. (1987). *Foundations of Cognitive Grammar. Vol. 1*. Stanford: Stanford University Press.

Langacker, R. (1991). *Foundations of Cognitive Grammar. Vol. 2*. Stanford: Stanford University Press.

Lakoff, G. (1987). *Women, Fire and Dangerous Things*. Chicago: University of Chicago Press.

Rosch, E. (1983). "Prototype classification and logical classification: The two systems" en E. Scholnick, *New Trends in Cognitive Representation: Challenges to Piaget's Theory*. Hillsdale, NJ: Lawrence Erlbaum Associates: 73-86.

Sweetser, E. (1990). "From Etymology to Pragmatics: The Mind-as-Body Metaphor" in *Semantic Structure and Semantic Change*. Cambridge: Cambridge University

Wittgenstein, L. (1953). *Philosophische Untersuchungen*. Traducción al inglés de G. E. M. Anscombe: Philosophical Investigations. Oxford: Blackwell.

EL IDEAL DE LA CIENCIA UNIFICADA Y SUS FUNDAMENTOS

M.Carmen Hernández Martín
Universidad de Sevilla

El ideal de unidad de la ciencia aparece ligado a la primera gran escuela contemporánea de filosofìa de la ciencia: el positivismo lógico. Como piedra angular del planteamiento positivista se inició la Enciclopedia de La Ciencia Unificada. Unidad de la ciencia, natural o social, en torno a un lenguaje común: el lenguaje de la lógica moderna. Herramienta principal, al menos en el enfoque de Carnap, es el reduccionismo a nivel de conceptos y a nivel de teorías.

La Enciclopedia de la Ciencia Unificada fué idea de Otto Neurath allá por los años 20 del pasado siglo; así lo testimonian Carnap y Morris en el primer volumen de *Foundations of the Unity of Science*[1]. Los dos volúmenes de esta obra iban a ser los primeros de la Enciclopedia que debería tener unos 26 en total. El primer volumen estaba terminado en 1939 y se proyectaba el segundo para 1944. Pero múltiples problemas, la guerra mundial entre ellos, retrasaron años su publicación, en la que el estilo había evolucionado tanto como para dar cabida entre sus monografías a la *Estructura de las Revoluciones Científicas* de Kuhn.

Como nos dice Carnap la muerte de Neurath en 1945 privó al movimiento de unidad de la ciencia de su ímpetu organizativo central. En efecto el proyecto total quedó interrumpido.

Podemos retroceder en este ideal de unidad de la ciencia en torno a un lenguaje, al Renacimiento, a los orígenes de la Revolución Científica, a esa frase de Leonardo, asumida después por Galileo: la naturaleza está escrita en lenguaje matemático. Aquí la unidad, más modesta, se refiere a las ciencias naturales, que efectivamente en los siglos posteriores han mostrado una tendencia creciente a la unificación en torno a la matematización. Matematización que plasma en una serie de rasgos, entre los que está la cuantificación, que no es siempre el más importante, ni siquiera imprescindible.

[1] Neurath, O., Carnap, R. & Morris, Ch. Foundations of the unity of science, Chicago, 1939-70.

La matematización puede ser perseguida hasta la antigüedad griega, hasta el pitagorismo y el atomismo, con sus consecuencias en la obra platónica y arquimediana. Encontramos un primer modelo explicativo matemático correspondiente a la música; siguen los modelos astronómicos, vinculados al anterior; por último asistimos al nacimiento de los primeros conceptos físico-matemáticos en la Escuela de Alejandría.

La esencia de esta matematización podría ser condensada en los siguientes rasgos:

a) el uso de conceptos de un grado de abstracción superior al físico. Los conceptos naturales ordinarios son sometidos a manipulación en cuanto a sus notas constitutivas. Por ejemplo, como ocurre en la física matematizada, concebir materia sin peso (en el principio de inercia) o peso sin materia (en el equilibrio de las figuras planas),

b) el planteamiento de situaciones ideales, incluso irreales, que nos permiten dividir los problemas concretos en diferentes aspectos, más fáciles de estudiar al ser así aislados, y que tendremos posteriormente que sintetizar a la hora de la aplicación de la teoría,

c) exactitud de las situaciones y modelos ideales creados, a los que la realidad accederá a menudo por aproximación

d) invención, incluso postulación de existencia, de entidades no observables, que situamos en los modelos ideales y englobamos en comportamientos exactos y en relaciones, de los que podamos derivar el nivel de lo observable,

e) extensión al límite de los significados de conceptos en principio observacionales, al ser aplicados a estas entidades postuladas,

f) cuantificación.

A estas dos posiciones, – unidad en torno a la lógica, unidad en torno a la matemática – muy vinculadas en el fondo, podemos añadir una tercera: la que basa la unidad científica en el empleo de un método esencialmente común, el método hipotético-deductivo. A pesar de la pluralidad de métodos específicos de cada disciplina científica, todas comparten la actividad de elaborar hipótesis, deducir de ellas consecuencias y someterlas a contrastación.

Este método da un paso más: una vez conseguidas y confirmadas una serie de hipótesis, las organiza según una construcción jerárquica. El método hipotético-deductivo culmina en el método axiomático.

También los orígenes del método axiomático están en Grecia y no sólo en los Elementos de Euclides, sino muy probablemente en todos los Elementos que se construyeron anteriormente y de los que sólo nos han llegado referencias. Y no sólo además en obras de pura matemática, sino en teorías y modelos de una cierta realidad física, como ocurre en los tratados arquimedianos sobre el equilibrio.

Y en esos tratados arquimedianos, nos dirá Galileo, se inspirará la revolución científica, que cristalizará finalmente en el sistema de Newton, quien al igual que Arquímedes va en busca de los "principios matemáticos de la filosofía natural".

En el positivismo del Círculo de Viena se unen las tres posiciones: el lenguaje de la lógica russelliano-hilbertiana es la herramienta al servicio de la matematización y del método hipotético-deductivo en su versión más fuerte. La lógica matemática o mejor, la lógico-matemática es al tiempo herramienta del análisis y modelo para la síntesis.

Las teorías son o deben ser sistemas axiomáticos. Los axiomas son las leyes de más alto nivel y a partir de ellos se deducen las leyes de nivel más bajo y todas las demás proposiciones.

Los términos teóricos, que aparecen en los axiomas, se justifican porque pueden ser definidos explicitamente a partir de términos observacionales. En seguida se ampliará este requisito a definiciones condicionales u oraciones reductivas para caracterizar a los términos disposicionales mediante las operaciones que hacen posible se les de un significado. "Conocemos el significado de un término si conocemos bajo que condiciones se nos permite aplicarlo en un caso concreto y en que condiciones no", nos dice Carnap. "Siempre que conocemos un método experimental de determinación de un término, estamos en situación de formular un enunciado de reducción para él"[2].

Es este el primer reduccionismo, el de los términos teóricos.

Años más tarde se nos dirá que las teorías son sistemas abstractos a los que se da un significado mediante un conjunto de

[2] Carnap, R. "Logical foundations of the unity of science" en Neurath, Carnap & Morris,1939-70, pags. 49,51

reglas de correspondencia. Se subrayará ahora la distinción dentro de los términos teóricos entre primitivos y definidos, con lo que los últimos recibirán su significado de los primeros y estos a su vez del contexto en el que están incluídos. El contexto habrá de ser consistente y de él derivarse en último término proposiciones que pueden ser contrastadas con la experiencia. Esta contrastación parcial dará su significado al sistema total. Los términos y las proposiciones aisladas no son los receptores adecuados del sentido. Se ha renunciado al reduccionismo fuerte de los primeros años, que no ha soportado la propia crítica interna.

Un segundo reduccionismo es el que se da entre teorías sucesivas. El desarrollo del conocimiento, la historia de la ciencia, se produce, según los positivistas, por acumulación. Y a veces una teoría, como la de Newton, sustituye a otra o a otras, como las de Galileo, Kepler y Huyghens, para abreviarlas y unificarlas. Las nuevas leyes se inducen de las viejas y estas últimas se deducen dentro del nuevo esquema teórico.

La crítica a este segundo reduccionismo vendrá en primer lugar de Popper, quien mantendrá que dificilmente pueden existir nexos deductivos o inductivos entre teorías que contienen proposiciones contradictorias. Y recurre justo a las relaciones Galileo-Newton y Kepler-Newton para mostrarlo.[3]

Un tercer reduccionismo es el que se da entre teorías simultáneas, es decir, entre diferentes ciencias.

En el primer tomo de la Enciclopedia[4], Carnap, tras distinguir entre ciencias formales y ciencias empíricas, divide a estas últimas en físicas y biológicas. Poco antes nos ha dicho que va a considerar ciencia todo conocimiento teórico, "pertenezca al campo de las ciencias naturales, las ciencias sociales o las llamadas humanidades; más aún se incluirá todo conocimiento basado en la aplicación de procedimientos científicos especiales, e incluso el conocimiento basado en el sentido común en la vida cotidiana", ya que "lo que normalmente llamamos ciencia es simplemente una continuación más sistemática de aquellas actividades que llevamos a cabo en la vida

[3] Popper, K.R. "The aim of science", Ratio, 1, 1957 b, pp. 24-35; reimpreso en Popper [1972]

[4] Carnap, R. "Logical foundations of the unity of science" en Neurath, Carnap & Morris, 1939-70, pags. 45-49.

cotidiana para conocer algo". Parece claro que los términos 'físico' y 'biológico' van más allá de su uso ordinario.

En efecto, " 'física' es un nombre común para todo el campo no biológico de la ciencia, englobando tanto las investigaciones históricas como sistemáticas dentro de este campo, incluyendo así, química, mineralogía, astronomía, geología etc."

La distinción entre física y biología descansará en la que existe entre organismos y no organismos. La biología se ocupará de los organismos. Y en sentido estricto biología será todo aquello que usualmente se considera así, esto es, biología general, botánica y la mayor parte de la zoología.

En sentido amplio la biología incluirá a la psicología y a parte de la fisiología (sistema nervioso) y de las humanidades, en cuanto estudio de organismos individuales y sus relaciones con el entorno; y a la sociología y a la mayor parte de las humanidades como estudio de grupos de organismos.

"La biología presupone a la física pero no a la inversa". "El biólogo tiene que conocer las leyes de la física para estudiar los procesos en los organismos. Las necesita para explicar esos procesos. Pero como no le bastan, le añade otras leyes no conocidas por el físico, es decir, las leyes biológicas".

"La cuestión de la unidad de la ciencia", nos dice Carnap, "es considerada aquí como un problema de la lógica de la ciencia, no de ontología". No nos preguntamos si el mundo es uno o si los problemas mentales son realmente físicos o espirituales, ni una serie de cuestiones de tipo metafísico discutidas por el monismo, el dualismo y el pluralismo. Nos preguntamos por "las relaciones lógicas entre los términos y las leyes de las diferentes ramas de la ciencia".[5]

Existe un lenguaje de la ciencia y unos sublenguajes: el lenguaje físico y el lenguaje biológico.

El lenguaje físico sólo contiene términos físicos, es decir aquellos que, unidos a los términos lógico-matemáticos, nos son necesarios para describir los procesos en la naturaleza inorgánica; aunque puedan aplicarse también a procesos en organismos. En él se enuncian las leyes físicas.

El lenguaje biológico, en cambio, contiene términos biológicos y físicos, unidos a los términos lógico-matemáticos. El lenguaje físico

[5] id. pag. 49.

es un sublenguaje del biológico. Pero las leyes y las proposiciones biológicas pertenecen al lenguaje biológico, no al físico.

Este último lenguaje está referido fundamentalmente a la biología en sentido estricto. No queda claro, en esta obra, el planteamiento de los lenguajes biológicos en sentido amplio.[6]

Pero es necesario distinguir entre el lenguaje de la ciencia en sentido estricto y un lenguaje-base cuyos términos pertenecen todavía al nivel precientífico. "Ese sublenguaje que es la parte común entre el lenguaje precientífico y el lenguaje físico puede llamarse lenguaje de la cosa física o brevemente *lenguaje-cosa*"[7]. Es el lenguaje que usamos para hablar de las propiedades directamente observables, aún no sometidas a medición. Añadimos aquellos términos que "expresan la disposición de una cosa a una cierta conducta bajo ciertas condiciones". Son los predicados disposicionales, reducibles a predicados observacionales ya que en base a estos podemos describir las condiciones experimentales y las reacciones que los caracterizan. A partir de aquí podemos introducir todos los restantes términos del lenguaje, por medio de definiciones y de oraciones reductivas.

"Podemos ver facilmente como todo término del *lenguaje físico* es reducible a los del lenguaje-cosa". El lenguaje-cosa proporciona la base adecuada para la física y para la biología. Afirmar que este lenguaje es base suficiente para cualquier rama de la ciencia, incluyendo a la biología en sentido amplio, constituye la tesis del *fisicalismo.*

Existe pues una unidad de la ciencia en cuanto a reducibilidad de términos y de lenguajes. Pero ello no implica la reducibilidad de las leyes. "En el presente estado de desarrollo de la ciencia no es posible ciertamente derivar las leyes biológicas de las físicas".[8] "Es obvio que en el momento actual las leyes de la psicología y de las ciencias sociales no pueden ser derivadas de las de la biología y la física". "Por otro lado no se conoce razón científica para asumir que tal derivación sea en principio y para siempre imposible". "No existe en la actualidad la *unidad de leyes*". Pero no podemos perder de vista este objetivo, aunque no sepamos si se alcanzará alguna vez. La unidad de lenguaje es la condición necesaria preliminar.

[6] id. pags. 46-48.
[7] id. pags. 52-60.
[8] id. pags. 60-62.

Es particularmente interesante la comparación de este Carnap de los años 30, en plena efervescencia reduccionista, con el Carnap de *Symbolic logic and its applications* de los años 50. En los sistemas axiomáticos construidos en esta última obra los términos específicos de cada ciencia aparecen como primitivos y en ningún momento se les trata de definir, ni siquiera condicionalmente, a partir del lenguaje más básico. En todo caso el concepto de definición que podría tratar de aplicarse sería el formalista hilbertiano de definición por los propios axiomas; con lo que, como ya había señalado Frege en su famosa carta a Hilbert, no se trataría de la definición de los términos, sino de la definición de la estructura en que están incluidos. Yo diría que Carnap ha superado ampliamente su propio reduccionismo.

Los planteamientos de unidad de la ciencia han sido muy criticados desde filosofías de las ciencias humanas y sociales. Recuérdese la famosa disputa con la Escuela de Frankfurt. Aquí se tiende a identificar con el positivismo a todo planteamiento filosófico que haya partido de las ciencias naturales. Se acusa a la ciencia natural de querer imponer sus métodos a las ciencias no naturales, ahogándolas. Se propone como método específico para las ciencias humanas el interpretativo o hermenéutico. Se contrapone al concepto de explicación, que se identifica con el concepto de explicación deductiva, el concepto de comprensión basado en la interpretación y anclado en gran medida en la empatía.

Así mismo encontramos críticas en el ámbito de los filósofos de la biología, especialmente de los biólogos evolucionistas; es el caso de Ernst Mayr.[9] También ellos se quejan de la presión por parte de planteamientos fisicalistas, desde los que sólo se reconocerían como científicos los métodos de la bioquímica o de la biología molecular.

Todas estas críticas, antirreduccionistas, ponen en evidencia el peligro de confundir unidad de la ciencia con uniformidad de la ciencia, creando un mecanismo represor en que unas ciencias nieguen el estatus de cientificidad a otras.

¿Existe una posibilidad de síntesis en que se respeten los métodos específicos de cada una de las ciencias - también las ciencias naturales se diferencian entre ellas - conservando no obstante aquello común a todas?

[9] Mayr, Ernst. 2004.

Podemos poner en duda si la matemática es el lenguaje adecuado para todo tipo de ciencia, o al menos si no es prematuro aplicarla en ciencias muy jóvenes o muy complejas. Podemos pensar que a veces no poseemos la herramienta lógica adecuada para ciertos ámbitos del saber, o al menos que no la poseemos todavía. Pero negar la posibilidad de unidad de la ciencia, como negar la posibilidad de unidad del conocimiento, al menos como ideal regulativo, es apostar por la incoherencia y por la irracionalidad.

Bibliografía

Carnap, R. (1934). *The unity of science*, London.

Carnap, R. (1937). *Logische Syntax der Sprache*, Wien, 1934. London.

Carnap, R. (1939-70). "Logical foundations of the unity of science" en *Neurath, Carnap & Morris*, Chicago.

Carnap, R. (1939-70). "Foundations of logic and mathematics", en *Neurath, Carnap & Morris*, Chicago.

Carnap, R. (1950). "Testability and meaning" en *Philosophy of Science*, 3, 1936, pp. 419-71, y 4, 1937, pp. 1-40. Reprinted: New Haven, Conn.

Carnap, R. (1954). *Introduction to Symbolic Logic and its applications*. New York, 1958. Berlín.

Hernández Martín, M.C. (1993). "Filosofías de la matemática y evolución del positivismo lógico", *Fragmentos de Filosofía*, 3, pp. 71-75, Sevilla.

Hernández Martín, M.C. (2001). "El análisis lógico del lenguaje científico. Los lenguajes-cosa de Carnap", en Nepomuceno, A., Quesada, J.F., Salguero, F. (eds.) *Información, tratamiento y representación*, pp. 217-229, Sevilla.

Mayr, Ernst. (2004). *What makes biology unique? Considerations on the autonomy of a scientific discipline*, Cambridge.

Neurath, O., Carnap, R. & Morris, Ch. (eds.) (1939-70). *Foundations of the unity of science,* Chicago.

Popper, K.R. (1972). "The aim of science", *Ratio*, 1, pp. 24-35, 1957; reimpreso en *Objective knowledge*, Oxford.

BEAUTY IN MATHEMATICS: AESTHETIC TERMS

Ulianov Montaño
University of Groningen, The Netherlands

Abstract. In this paper I will discuss the interpretation of aesthetic terms and descriptions in mathematics. I will show that Gian-Carlo Rota's and James McAllister's approaches to this issue are based on a reinterpretation of aesthetic descriptions. Then I will sketch a notion of aesthetic term that will allow us to give a literal interpretation of aesthetic terms and description in mathematics. This discussion is part of a longer research on beauty in mathematics, the approach that I am presenting here is still fragmentary but I put it forward expecting to arouse interest a further debate on the subject.

Aesthetics and Mathematics

Discussions and references about aesthetic qualities in mathematics are part of the language that mathematicians use every day, both in professional discussions and teaching. Remarks about the elegance of certain proof or the beauty of certain theorem are very common. If one pay close attention and take this kind of judgments seriously one will soon face two problems.

First, it is not clear what mathematicians mean when they use the term "beautiful" or other closely related terms. We are familiar with the use of aesthetic terms in every day life, but when a mathematician make a judgment about the beauty of a theorem we get the impression that his use of aesthetic terms is not exactly the same as the one in our daily life. It would seem that aesthetic experience for mathematicians is of a special kind, or, at least, it is not clear whether aesthetic experiences and language in mathematics is the same as in ordinary life.

The second problem is that the use of terms like "beautiful" or "elegant" denotes a preference for these terms as opposed to certain others like "true", "consistent" or "efficient". This preference mirrors the use of those terms in ordinary language: "truth" seems to refer to

objective, public matters while "beauty" has to do with subjective, private issues; "beauty is in the eye of the beholder" makes explicit that aesthetic judgments are a matter of opinion. Thus, being mathematics a science, concerned mainly with truth and objective matters, does the presence of aesthetic judgments introduces a tension regarding the nature of mathematics a science, as a rational activity?

One way of avoiding these problems is by explaining that aesthetic judgments in mathematics (and in science in general) must be re-interpreted in terms of the precepts, methods and aims of the discipline (mathematics as well as other science) and then showing how this new interpretation is related to the ordinary interpretation.

Here I propose another one of avoiding these problems. I will take aesthetic language in mathematics at face value, just as we would do with any other aesthetic description in ordinary language. Regarding the first problem I believe that the oddness of the mathematical use of aesthetic terms is not different from the oddness we would find in aesthetic judgments when applied to avant-garde art, for example. In both cases specific cognition process, knowledge and awareness of the subject of the work are necessary to aesthetically appreciate the work. My strategy to face the second problem will be to claim that aesthetic judgments are not completely subjective and that precepts in mathematics, as with most disciplines, are subjected to historical change and cultural modulation. Unfortunately, a deeper discussion of these problems would take more space that the one intended here, so I will discuss only the problem of giving a literal interpretation of aesthetics in mathematics.

If we take a *face value* or *literal* interpretation of aesthetic judgments in mathematics we must consider them as regular aesthetic judgments. This literal interpretation entails two things. First, that mathematical aesthetic judgment can be viewed under the light of full blown philosophical theories of art and the aesthetic, and not only resorting to pre-theoretical interpretations. Second, such an approach can be considered as a project in aesthetics, as well as a project in philosophy of mathematics. Thus, discussions and revisions of concepts can be expected concerning both fields. In particular discussions concerning standard approaches in aesthetics can be seen as attempts to broaden the reach of aesthetics, rather than re-interpreting mathematical aesthetic language.

Now, let us further clarify the difference between "literal" and "non-literal" interpretation of aesthetics in mathematics.

Gian-Carlo Rota in his 1997 paper "The Phenomenology of Mathematical Beauty" presents an insightful view of how beauty in mathematics can be interpreted:

> "The term "mathematical beauty" together with the light-bulb mistake, are tricks that mathematicians have devised to avoid facing up to the messy phenomenon of enlightenment. The comfortable one-shot but misleading idea of mathematical beauty saves us from having to deal with the messy situation of a concept having degrees. All talk of mathematical beauty is a copout from confronting the logic of enlightenment, a copout that is intended to keep our formal description of mathematics as close as possible to the description of a mechanism. This copout is a step in a cherished activity of mathematicians, that of building a perfect world immune from the messiness of the ordinary world, a world where what we think should be true turns out to be forever true, a world that is to be kept free from the disappointments, the ambiguities, the failures of that other world in which we are forced to live." (Rota, 1997)

A better attempt to explain beauty in a science is presented by James McAllister in his 1996 book "Beauty and Revolution in Science". He develops the idea that aesthetic judgments in science are governed by a set of preference criteria for properties of theories, sets which he calls "aesthetic cannons". A particular property P of a theory has an associated weight W which assesses the preference for P in a specific scientific community. The aesthetic canon of that community is the (possibly infinite) set of pairs (P, W) that the community holds. According to McAllister the preference of the community for a particular property will increase if that property is often present in empirically successful theories. Thus the weight W of property P becomes higher if P is associated with successful theories. As the time passes and property P accumulates a record of presence in more successful theories, scientists tend to prefer theories with property P more than theories lacking P. That is, they start to develop a "taste" for P due to its association with empirical successful theories. McAllister calls this phenomenon *Aesthetic Induction*. Aesthetic descriptions in science are expressions of scientists' preferences, but

another, more important, expression is their tendency to choose among rival theories on the grounds of their developed taste.

McAllister fills the gap between ordinary aesthetics and his inductively governed aesthetic canons by showing that we can establish analogies that link taste in one and the other side. His examples are particularly sound when it come to applied art. For example McAllister shows that steel structures, which were increasingly used in buildings due to their structural success, eventually developed an appreciation among architects and the public. Thus, a property that appeared in buildings due to their structural success later gained a preference, that is people started to appreciate them on the grounds of taste and not only because they found them useful. This example shows remarkable resemblance with the development of taste in science due to the frequent appearance of certain properties in empirically successful theories.

Non-Literal Aesthetics in Mathematics

In the following, we will see how two literal interpretations of aesthetic terms in technical aesthetics seems to contradict the Rota's and McAllister's approaches, and prevent thus a complete interpretation of these approaches as face value aesthetics views.

Isabel C. Hungerland has proposed that aesthetics terms must be approached focusing on their linguistic and logic properties. I will refer to Hungerlan's view as the logic-term approach. Peter Kivy, attending to some problems with Hungerland's approach, has proposed an alternative which will be referred here as the terminal approach.

The logic-term approach usually needs a supplementary view of what a term refers, so it is usually coupled with a strong view on aesthetic phenomena. The non-utilitarian view of aesthetic phenomena will help us to make clear how Rota's and McAllister's approaches (which see aesthetic descriptions as indicators of enlightenment or empirical success) are not face value interpretations of aesthetics in mathematics.

According to the non-utilitarian view aesthetic phenomena are characterized by being instances of *disinterested attention*. In aesthetic contemplation the attention of the beholder is directed to some object

(concrete or not) with no ulterior purpose; the beholder gains nothing else and intends nothing else but the contemplation in itself. In this sense, the attention is disinterested; there is no external interest driving the attention. According to this view attention can be present in two ways: interested or disinterested. Interested attention is an element of an action leading to an external aim, an aim different from the object of attention itself. Disinterested attention has no further aim but the object of attention itself. Paying attention to a property of a theory can be done, according to this view, in two ways: Interested or disinterested. Interested attention is involved, for example, when we focus our attention on certain properties of two rival theories in order to choose or evaluate them. Choosing a theory, in this case, is the final aim; we focus our attention to pick a theory or to evaluate it in comparison with another. In an episode of disinterested attention, in contrast, a scientist would direct his attention to that property with no other aim but the contemplation in itself. The interpretations proposed by Rota and McAllister both suppose a further aim, enlightenment and theory choice respectively. So they seem to be of the kind of interested attention.

The non-utilitarian view as a theory of aesthetic phenomena has its own problems; however, there are more reasons to think that McAllister's and Rota's views are non-literal. In order to show this let us complete the logic-term approach.

The non-utilitarian view is a theory about aesthetic phenomena; the logic-term approach is a complementary theory about the characteristics of aesthetic terms. According to this approach once established the phenomena to which aesthetic terms refers, it is necessary to understand the "logic" of the terms. In this respect there are two things to address: first, the truth conditions that govern the use of terms in descriptions, and, second, the application-criteria that justify the utterance of aesthetic descriptions.

According to Isabel Hungerland, aesthetic terms refer to properties and they are part of a distinctive group that differs from other "non-aesthetic" terms. An example of a non-aesthetic term is the term 'strong' and an example of an aesthetic term is 'elegant'. Hungerland clarifies how aesthetic terms, which she calls A-terms, by showing that they are governed in a different way than non-aesthetic terms, which she calls N-terms.

Non-aesthetic and aesthetic terms differ in the way in which they are used in descriptions in the following aspect:

N-terms, non-aesthetic terms, can appear in sentences of the following two types:

(1) 'John is strong'
and
(2) 'John looks strong but he is not'

A-terms, aesthetic terms, *cannot* be used in the way indicated in (2). There is little or no difference for aesthetic term to be use in either way 'John is elegant tonight' or 'John looks elegant tonight'.

Furthermore, type (2) sentences like 'John looks elegant but he is not' seem to make little or no sense.

For any non-aesthetic term N there is a difference between being and looking N. Something can look N but it still might be not N. In contrast, for aesthetic terms there is no such thing as an *is-looks* distinction.

This is so, Hungerland argues, because aesthetic terms are terms devised precisely to talk about how things look, that is, the criteria of application of aesthetic terms depend entirely on the subjective experience of the subject who make use of them. While in the use or application of a non-aesthetic term like 'strong' there are several external or objective criteria that can be use to correct the initial use of such a term. For example, let us suppose that we have just said that John was strong. Then some one can tell us that John is very ill or we can corroborate ourselves that John is a weak person. We have external reasons that allow us to correct from 'John is strong' to 'John looks strong but he is not.' That is so because both of these sentences are governed by external objective criteria.

In contrast, there are not external or objective criteria to correct the use of aesthetic terms. If John looks elegant to us and we claim that John is elegant, no external criteria can induce us to then claim that John looks elegant but he is not. This last claim simply makes no sense.

Now, attending to these considerations, let us come back to aesthetic descriptions in mathematics.

Rota's view on aesthetic descriptions in mathematics, mathematical beauty is a trick to conceal the phenomenon of

enlightenment, fails to avoid the *is-looks* distinction. For example, a mathematical proof can be qualified as beautiful, in Rota' s sense we must think that this judgment amounts to qualify the proof as enlightening, however due to the characteristic fuzziness of enlightenment, as Rota himself proposes, a proof can later be shown to provide no enlightenment at all. Rota himself provides examples of how some fields of mathematics can stop being considered beautiful. For example the field of mathematics of which the proof is part can become out of fashion; that was the case of analytic projective geometry. In this case we have a situation in which a proof looks (or looked) enlightening but is not really enlightening (any longer). The interpretation in terms of enlightenment brings along an *is-looks* distinction, which prevents the use of terms like 'beauty' from being face value aesthetic terms according to the logic-term approach presented above.

McAllister approach faces a similar problem. Lets us suppose that the property of possessing a mechanical model is a property highly valued according to our aesthetic canon. However, our judgments concerning the possession of a mechanical model can (and as a matter of fact, they must) be corrected using external or objective criteria. In order to claim that the theory T posses a mechanical model one must show that model or show a method to construct one. In fact, the very mechanism that governs the development of an aesthetic canon is an external mechanism, because it depends on the empirical success of the theories that possesses the property. In the general case the use of aesthetic terms governed by aesthetic induction is, in time, governed by sub-criteria that are bound to be external because aesthetic induction depends on empirical success. Thus, McAllister interpretation of aesthetic judgments in science fails to meet the characteristics specific to aesthetic terms under the term-logic approach.

The term-logic approach is not free from problems and has been notoriously challenged by aestheticians like Peter Kivy. Kivy has proposed an alternative view, which I will refer to as the terminal approach. One of the problems with the term-logic approach is that it split terms in two categories; A-terms and N-terms. However, when we make aesthetic descriptions, we very often use N-terms. We can describe certain musical progression as 'strong', or certain picture as 'balanced', or the work of an artist as 'solid'. It seems that what

makes a term aesthetic is not the inherent logic and linguistic criteria that governs their use or justify our utterance of them. Rather, Kivy proposes that a more accurate view can be achieved if we look at the characteristic role of aesthetic descriptions rather than at the logic-linguistic features of the terms. Kivy argues that aesthetical terms are "terminal", in Kivy's words:

> "aesthetic terms, then are those that, characteristically, occur in descriptions which seem to be ends rather than beginnings… they are terms that do not provide the reason for anything else." (Kivy, 1975)

According to this approach, there is no specific set of terms that can be labelled as "aesthetic"; rather the "aesthetic" quality of terms comes from playing a role in aesthetic descriptions. The division between non-aesthetic N-terms and aesthetic A-terms was supposed to be based on the existence of an *is-looks* distinction for N-terms and the lack of it for A-terms. However, as we have seen, N-terms can be used in aesthetic descriptions.

For Kivy the contrast is not between N-terms and A-terms. Nor between the presence and absence of an *is-looks* distinction, but rather between non-terminal and terminal terms: Non-aesthetic terms, when used in a description, serve as a "prelude to something else: the premise of an argument or a call to action". Instead of resorting to the *is-looks* distinction, Kivy makes clear the distinction between non terminal and terminal terms by establishing the contrast between moral and aesthetic terms to make clear what he mean by "terminal".

> "…the fact that aesthetic descriptions are "terminal", that they lead nowhere, distinguishes them sharply from moral descriptions, which often are preludes to action. To conclude that a course of action is right is to provide some reason for pursuing that course of action in the future. To conclude that a man is greedy is greedy is to provide some reason for future actions, or attitudes towards that man. To conclude that a novel is "unified," a painting "garish," a poem "sentimental," a symphonic movement "sad," however, provides no reason for anything except continued contemplation, or an end to it." (Kivy, 1975)

Kivy claims that a terminal approach to aesthetic terms points to the right direction, however it does not provide a complete answer.

Now, I believe that even with this incomplete answer, Rota's and McAllister's proposals face difficulties. I will evaluate them here based on the ideas behind the terminal approach.

Rota's "enlightenment" approach sees aesthetic descriptions as masqueraded descriptions concerning the enlightenment provided by a proof, for example. In this sense they can be construed as descriptions about the enlightening provided by of the proof. To provide understanding is clearly a virtue, but, what are the reason to make such descriptions?

Let us think of sentences like 'P is a correct proof', 'T is a consistent theory', or 'S is a true theorem'. All these sentences can naturally lead to a further conclusion or a future action. For example, we can use theorem S to prove another theorem, and then 'S is a true theorem' can be part of our argument.

Now, sentences like 'P is a correct proof', 'T is a consistent theory', or 'S is a true theorem' are similar in the sense of being descriptions concerning epistemic virtues. In this sense, descriptions that qualify the ability of a proof to provide understanding are similar to descriptions concerning other epistemic virtues of a proof. Thus, we have no reason to believe that descriptions like 'P is enlightening' are different in this respect: they are also about an epistemic virtue. And they thus can lead to further conclusion.

Under Rota's interpretation, aesthetic descriptions in mathematics would not be terminal and they would be better treated as the rest of epistemic descriptions in mathematics.

McAllister view fares better, however, it still cannot be viewed as a face value approach to aesthetics in science. His view is based on the idea that a set of preferences, an aesthetic canon, drives scientist to put forward aesthetic descriptions. It is true that, being governed primarily by preferences, such descriptions cannot be used to lead to a further conclusion, that is, they cannot be assimilated to epistemic descriptions. However, as the hierarchy of preferences in the aesthetic cannon is determined by the past empirical success of specific properties. Let us remember that there is a series of weight associated to properties that assesses the intensity of the preference and allows us to compare among the different properties of the canon. The use of an aesthetic description is not only the result of a process of aesthetic

induction, but also a part in the process of increasing the preference for future appearances of the term involved in the description. This is apparent when scientists decide to choose one theory rather than another based on aesthetic criteria. In McAllister's view, theory choice is the natural result of aesthetic judgment. In this sense, McAllister's interpretation is similar to the moral descriptions which Kivy uses to exemplify non-terminal terms. To conclude that a theory T is beautiful is to confirm its preference by the aesthetic cannon. This, in time, due to the inductive process behind it**,** *which makes sense only if we are comparing theories*, can naturally lead to prefer a theory over another. To conclude that a theory is beautiful, then is to provide a reason for pursuing a course of action in the future, for making a choice between theories. And even if no choice is made, the future attitudes towards theories would be altered. In this respect McAllister's aesthetic descriptions in science are analogous to the moral descriptions as pictured by Kivy. Thus, McAllister's approach is not terminal.

Rota and McAllister's approaches to beauty in mathematics and science certainly give valuable insights. In order to make a fair assessment of their achievements, we must remember that they are not trying to put forward the basics of aesthetics in mathematics. Rota is interested in the relevant epistemic features related to the phenomenology of aesthetic descriptions in mathematical practice. McAllister's project is a defence of a rationalist image of science; he articulates a model of scientific change that is consistent with a rationalist view of science.

Literal Aesthetics in Mathematics

Now, my project is mainly concerned with a face value interpretation of aesthetics in mathematics. I still cannot provide a complete answer, but I can put forward the elements that, I believe, point to the right direction.

Literal approaches to aesthetic terms are not free from drawbacks. Term-logic approaches suppose that there is a specific set of aesthetic terms, but we have seen that non-aesthetic terms frequently appear in aesthetic descriptions.

Kivy's approach characterize aesthetic terms as being terminal, however art-criticism, the paramount of aesthetic description, uses descriptions aimed to do something else than merely presenting a terminal remark. According to William Hyde (Hyde, 1978), art criticism descriptions, by explaining how certain work of art relates to a wider art-world, makes a work of art available. Kivy himself recognize that his view is only a partial view, so here I will not discuss in detail this problem.

Now, what I will try to do in the following is to present my view on how to understand aesthetic terms in order to encompass aesthetic judgment in mathematics. I agree that aesthetic terms are characteristically terminal, but I must also add that this "terminality" is the result of a background process. My proposal is to consider aesthetic phenomena as a process and aesthetic terms as participants in the final sub-process of that process. The characteristic terminality of aesthetic terms is subject to internal modulation by the aesthetic-process in which they participate. Aesthetic terms participate in an aesthetic-process as a terminal stage and as such their terminality depends on the previous stages of the process.

Now, in order to further explain my approach I need to address some problems related with the aesthetic phenomena.

First, I believe that the notion of form in aesthetics will provide the grounds to understand the kind of object of aesthetic judgment. Form, in my approach, would serve as object and source of the aesthetic experience.

Second, I must explain the subjective internal events engaged in an aesthetic experience. In this respect here I will only address the most basic kind of events phenomena, emotional phenomena, and I will left out more complex formal events. Concerning emotions, I will resort to a model of emotion endorsed by Joseph LeDoux and Jenefer Robinson which gave the inspiration for my own process model of aesthetic phenomena.

Form and Formalism

Aesthetic formalism was first proposed by the German musicologist Eduard Hanslick in 1854, as a response to Romantic expressionist aesthetics. Music was at that time a serious challenge for

traditional aesthetics. During the 20th century, with the arousal of non-representative painting, formalism became important in other artistic disciplines. As musical formalism is particularly well suited to deal with aesthetic phenomena in mathematics, due to the similarities between mathematics and music, I will focus on musical formalism, which is the most robust and longer developed kind of aesthetic formalism. One of the drawbacks of formalism is that the notion of form is ill defined. However, a provisional notion of form in music can be developed from the implicit notions used by traditional and contemporary formalist approaches. I will not provide he the details that led to the notion of form that I will present here, I will only claim:

> **(1)** that form is both the main subject and the main source of aesthetic phenomena
> and
> **(2)** that a notion of form as an abstract object can be can be implicitly defined by a set of precepts.

The following is the result of my analysis of form as it will be presented in my doctoral dissertation:

There are two things that must be noticed. First, in devising it I have restricted myself to the case of traditional music. Second, I have neglected here the problem of the emotion response associated to forms. That will be the subject of another section below. So, here I deal mainly with the cognitive features of form.

A Musical Form fits the following description:

A) Non-concreteness

> **A.1)** Musical Forms are abstract structures. They are not physical events, like sound; they have no causal dependence or space-temporal location. They are not instances of music itself, i.e. they are not acousmatic events as proposed by Scruton.
> **A.2)** In analogy to musical events that take place in an acousmatic space, musical forms are events that take place in a phenomenological cause-independent and instance-independent space. I have taken the term phenomenological from Scruton but

I will give it only a restricted use; by phenomenological I refer only to the quality of events that are experienced by the listener; in contrast to the experiences of the acousmatic space, that are musical experiences, the experiences that I am referring to are 'cognitive' experiences.

A.3) The general properties of musical forms, thus, will be determined by the kind of events that are allowed to take place in this phenomenological space. We have to establish, at least informally, the features of this space.

B) *Formal Scene*: Internal Syntax

Musical Forms are events that take place in a "space" analogous to the acousmatic space. I will not give a formal characterization of this "space", and in order to avoid confusion with the technical meaning of the term *space* I will use the term *scene*. Thus, my analogue to the *acousmatic space* could be called *formal scene*. I will informally characterize a formal scene as follows:

B.1) The formal scene is a 'cognitive' analogue to the acousmatic space, but it differs from a phenomenological auditory-space in three respects:

B.1.1) It exhibits physical causality independency and instance independency, but it is culturally dependent; furthermore, it is culturally and historically determined. The formal scene is determined by a set of norms (several syntactical systems for rhythm, harmony and modulations) that vary according to cultural standards. By contrast, the proposed causal-independency and instance-independency are not determined by cultural standards but by perceptual mechanism that allow us to recognize music and patterns in the music. In other words, experiencing music is something that has to do with the way that a normal perceptual and cognitive system in human beings determine, but experiencing musical forms is the result of an active cognitive processing of the experienced musical material. The adaptative nature of human cognition is responsible for the plastic, cultural and historical character of musical forms.

B.1.1.1) The cultural standards for notions like harmony, tension, rest or resolution are the norms or, if you want the 'syntactical systems' , that organize, that give an order to, the formal scene. That leads us to:

B.1.2) Drawing an analogy with the idea of *ordered space* we could say that the formal scene is an "ordered scene" with at least two "dimensions". (The number of dimensions is also culturally determined; the two chosen dimensions presented here correspond to the traditional harmonic system of music). One of the dimensions, the "horizontal" dimension is ordered according to the syntactic rules of rhythm. The other, the "vertical" dimension, is ordered according to the syntactic rules of scales, chords, and modes.

B.1.3) In addition to the two (or more) dimensions and their respective syntactical systems, there is an inter-dependence between them, governed by the rules of chord progressions and tonal modulation.

C) External Syntaxes

In addition to the 'internal syntaxes' that organizes the basic dimensions of the formal scene; there are 'external syntaxes' that allow 'constructing objects' in the formal scene. I don't think that we have to take the idea of 'constructing objects' literally, rather we can see the 'external syntaxes' as rules that govern the introduction of events in the scene. Given that the forms are events in the formal scene, we can use the metaphor of building objects just as a short way of naming "a proper introduction of events in the formal scene". The two methods that we can use to construct musical forms (to introduce proper events in the formal scene) are what I call *substructure-embedding* and *referentiality-embedding*.

C.1) Substructure-Embedding

Substructure-embedding refers to the embedding of structures already present in the scene to build new forms.

If there is no embedding in a form we will call it a First Order Form. If there is a single embedding; Second Order Form. And

if there are two embeddings; Third Order Form. First order forms corresponds to what I called form-3 (the basic harmonic relations of musical material); second order forms corresponds to form-2 (the arrangements of motifs, melodies etc); and third order forms corresponds to form-1 (the arrangements of whole works like symphonies). The structure of symphonies or concerts and the arrangements of motifs inside their sections were historically developed. Harmony seems to be more stable. That is the reason why first order forms differ from second and third order forms. And the historicity of form-1 and form-2 is the reason why traditionally determined embedding procedures are preferred to recursive embedding.
The way embeddings take place obeys different rules for Second and Third Order. Although recursive embedding is possible, the traditionally established patterns of embedding will be preferred. The embedding is governed by culture-dependent "syntaxes".

C.1.1) For first order forms we have the syntaxes of harmony and rhythm, already present in the structure of the formal scene. Following those syntaxes does not result in embedding, it only determines whether a certain structure is a musical form or not.
C.1.2) For second order forms we can embed only first order forms; we cannot use any musical event that is not already a first order form, like a note or a chord, as part of the syntactical structure; and the embedding must follow the rules of traditional musical forms like binary forms, sonata, rondo etc. The arrangement of those forms determines the 'syntax' for second order substructure-embedding.
C.1.3) For third order forms we can embed only second order forms, we cannot use any basic musical event (notes, chords, etc.) or any first order form as part of the structure; and the embedding must follow the rules of the major musical forms, like symphonies or concertos. The arrangement of those forms determines the 'syntax' for third order substructure-embedding.

C.2) Referentiality-Embedding
Referentiality-embedding refers to the inclusion of non-formal properties in musical structures. The non-formal properties to be included must fulfill two conditions:

C.2.1) First, they must possess referentiality (otherwise they could be treated as formal properties). Second, they must play a role in the constitution of the aesthetic value of the form in which they are being embedded (I have to recall that I am dealing only with the cognitive aspect of form, the most basic 'empirical' issues will be addressed in the next chapter).
Emotion-qualities and basic-beauty (properties of events involved in basic-type explanations) properties are examples of properties that could be introduced as part of musical forms through referientality-embedding.
C.2.2) The incorporation of these properties as part of a formal structure can be done in two ways:

C.2.2.1) First, if the inclusion of the non-formal property P is the result of a process of substructure-embedding that has the quality of being conservative of P. For example, if embedding a *beautiful* first order motif **A,** results in a *beautiful* second order binary form **AB**, we will say that the property of being beautiful has been preserved by the embedding, and that this embedding is beautiful-conservative. In this example the embedding of the first order form **A** to produce the second order form **AB** preserved the property of being beautiful while passing from **A** to **AB**.
C.2.2.2) Second, if inclusion of the non-formal property P is necessary to complete a certain pattern. Patterns, in this context could be understood as generalizations of syntaxes. I will focus only on the simplest and most widely used pattern: resolution. For example, if we already have two sections (second order forms) **A** and **B** arranged as **AB** and **A** is a restful section and **B** a tense one, the resolution pattern needs a rest section in order to obtain an entire emotional-resolution pattern. In this way, we are allowed to add a third *restful* (as it is asked by the

resolution pattern) section **C**. We obtain the form **ABC** that completes the resolution pattern *rest-tension-rest*. Notice that **C** is not necessarily different from **A**. If we use **A** to complete the pattern we obtain **ABA**; the traditional sonata form.

Remarks

Musical forms are events occurring in the formal-scene I have sketched. They are abstract structures only in the sense that they are cause and instance independent and that they have an internal order. I think that their ontological status is more akin to a fictive narrative than to other abstract objects, like platonic mathematical objects. In other words, musical forms are not necessarily abstract *objects* in the sense of objects actually existing with no spatial-temporal location. In this proposal we do not need to endorse Platonism in music; we only need sets of syntactical rules and subjective properties. Both of them (syntactical rules and subjective properties) could be construed as fictions, in the same way as we construe statements in a fictive narration (this is important, given that mathematical theories could be construed as fictions, as well).

This idea of form in music can be adapted to encompass mathematical results and processes as subjects of aesthetic judgment.

Emotions

Once we have an idea of what is the subject and the source of aesthetic experience, the natural complement to understand aesthetic phenomena is a description of the events involved in them. There are several ways to describe aesthetic experience; the non utilitarian view presented above is one of them. Here, I will concentrate only in a basic way of looking at aesthetic experience: the emotional response associated to it.

Emotions are the kind of phenomena that must be addressed in order to improve our understanding of what, in my analysis of form, I have called "basic" aesthetic events.

Emotion is not only the event associated to aesthetic experience, but it is often the primary response, and, in many cases, the

component responsible for our primary recognition of an experience as an aesthetic experience. In addition, the *pleasure* afforded by aesthetic contemplation is often the reason that makes art valuable, that is also the case with mathematics. One of the reasons why aesthetic descriptions are used in mathematics is to point out that some result or process in mathematics is able to arouse a pleasurable feeling once one get to grasp its full meaning. This component, the emotive component, has been neglected to some extent by Rota and McAllister (however, Theo Kuipers tries to put the psychological component back in the picture in Kuipers 2002).

The discussion about emotions is a long one. Here I will summarize Jenefer Robinson's emotion theory, which is a theory devised in order to deal with the problem of expression of emotions in art. The key element in this theory is that emotions are seen as processes instead of just physiological or psychological states of an individual. The notion of emotion as a process was introduced by the psychologist Joseph LeDoux during the 1990's.

Robinson's theory of emotion proposes that emotions are processes in which a special kind of automatic 'affective appraisal' induces physiological and behavioural changes which are then followed by what she calls 'cognitive monitoring' of the situation; this last component in turn interacts with the other components of the process. The emotion-process itself is always initiated by an automatic affective appraisal that occurs prior to reflection but the cognitive component is present in many aspects of the process and it can even take the place of the external environmental stimuli that are being assessed by the affective appraisal component. This is one of the important points of Robinson's theory: even if the process starts with an automatic appraisal, the stimulus that triggers it can be cognitive in nature. All these facts would explain some of the traditional paradoxes of emotions, like the fact that emotions are in some ways immune to assessment as rational or irrational.

There are several features that empirical and philosophical research on emotions has made clear. Firstly, emotions are not just judgments; however some kind of appraisal or evaluation seems to be necessary for emotion. Secondly, there is a connection between emotions and our interests, wants, wishes, values, and goals. Thirdly, there is a disagreement about what kind of evaluation is necessary for emotion: evaluations in emotions can be embodied in a belief, an

unconscious appraisal, a merely thought-of, a perception or seeing-as, or in something that looks like a reflex action. Finally, a general problem is the lack of an account of the psychological and physiological changes associated with emotions.

Experimental psychology and neurophysiology have shown that that affection can precede cognition in emotion, and this finding agrees with some traditional "puzzles" of emotions that some theorists have recognized. For example, the propositional content of an emotion need not to be believed; it can only be held in mind. Another example is that emotions are resistant to summing up and qualification; and, finally why we can have inconsistent beliefs when we experience mixed feelings.

One problem of the results of experimental psychology is that they do not allow us to understand emotional responses to complex cognitive stimulus, like the emotional responses to reactions to moral or economic or political beliefs. I can get angry at the fact that the president keeps making stupid decisions.

Joseph LeDoux's model of emotions pictures them as processes triggered by an affective evaluation, an non cognitive appraisal, of the environment, which in time goes trough two different systems in the brain, one related to basic emotive responses and another related to higher cognitive functions. This model suggests a way to integrate the puzzles presented by emotions mentioned above: The affective evaluation generates autonomic and motor changes, action tendencies and behaviour. All this activity is also canalized to the cortex where it is further cognitively processed.

Based on this idea Robinson tailors an emotion theory intended to be used later in her explanation of expression in art.

Robinson presents her description of an *emotion process* as follows:

> "…there is (1) an initial affective appraisal of the situation that focuses attention on its significance to the organism and causes (2) physiological responses of various sorts –specially ANS [autonomous nervous system] activity and changes in the facial musculature– and motor responses, which get the organism dealing with the situation as very broadly appraised by the affective appraisal, and which gives way to (3) a further more discriminating cognitive appraisal or monitoring of the

> situation. In other words, emotion is not a feeling, or a response, or a mental state, or a disposition; emotion is a process, a sequence of events. An affective appraisal draws attention to something in the environment significant to me or mine and gets my body ready for appropriate action. Then immediately cognitive evaluation kicks in, checks the affective appraisal to see if it is appropriate, modifies autonomic activity, and monitors behaviour. More complex cases of emotion in human beings might involve affective responses not to a perception but to a thought or belief, and the cognitive monitoring may be correspondingly sophisticated, but at the core of emotion will always be physiological responses caused by an autonomic affective appraisal and followed by cognitive monitoring."

Robinson claims that a more primitive version, one where cognitive monitoring is replaced by another (monitoring) function, is at work amongst lower organism. In that way Robinson gives a unified theory for the empirical results and the refinements of judgment theories. Cognition is responsible for the complexity of emotional arousal (emotions can be triggered by external stimulus as well as internal beliefs), the nuances (embarrassment is closely related to shame, for example) and plasticity (we can acquire emotional responses and then we can re-learn them) of human emotions.

An important aspect of this theory is that it involves several sub-processes identified by experimental psychologists. The exact sequence of these sub-processes is not fixed and there is evidence that the sub-processes overlap and that they are relatively independent of each other. For example, the hormonal response triggered by fear or anger could leave hormones that remain in the blood for hours or days, even if our cognitive monitoring tells us that there is nothing to be afraid of or angry at.

In addition to incorporating the most salient aspects of emotions studies by psychologists and philosophers, Robinson's theory can incorporate the effect of naming and labelling an emotion. In order to understand labelling emotions we must understand the general role of higher cognitions in emotion, especially their role in triggering emotions.

Cataloguing Emotions

Robinson argues that one of the important ways of dealing with emotions in everyday life is by attempting to explain our physiological state or behaviour by means of naming our emotional episodes. This fact is important for Robinson's views on art; as we will see later, for Robinson naming emotions shares the same principles and motivations as expression in art: both are attempts to elucidate emotions; the difference being that art is capable of a better articulation and individualization in this respect.

A second important fact is that the labelling of emotions can affect the development of emotions themselves. For example, thinking that I am sorrow instead of ashamed will lead to a sorrow behaviour and this, in turn, will feedback into the emotion process, resulting in an actual change in the feeling and other components of the process; eventually all these will lead to a change in the emotion itself. This is a fact that has been corroborated empirically.

There is evidence that supports the claim that our posture, vocal and facial expressions, and action tendencies tend to change our emotion-physiology and the way we feel. The research of the psychologist Paul Ekman has shown that changing the posture of facial muscles can induce changes in the physiological states and in the self reports of subjects (reports about the subject' s emotional state). In the same way, naming an emotion (naming it "sorrow", for example) may induce the behaviour and bodily states appropriate to that emotion.

This shows that cognitions as naming or labelling an emotion can modify the emotion-process; the way in which we 'catalogue' our emotional responses influences the development of the emotion process itself. Thus, thoughts and beliefs and other 'ways of seeing situations' can figure in the causal chain that leads to an affective appraisal.

Robinson speculates that in the cases of 'cognitively triggered emotions', my being angry at the stupidity of the president, for example) there must be a meta-appraisal that 'evaluates' or appraises the thoughts or beliefs in a similar way as we would appraise an environmental situation. (Robinson, 2005, p.91)

Aesthetic Phenomena as Processes

LeDoux' and Robinson's ideas have been useful to accomplish their own aims. Thus, it might be useful to import their idea of picturing internal events as process. We will see that it looks appealing to cope with the problem of aesthetic terms. Although the approaches of Hungerland and Kivy are elucidating we must remember that we are facing the problem of presenting a view on aesthetic terms that is both coherent with face value aesthetic interpretation and capable of accounting for the use of aesthetic judgments in mathematics.

My proposal is thus to consider aesthetic terms as tags, analogous to emotion tags: aesthetic appreciation phenomena should be considered a process containing several sub-process stages. We will consider as a sub-process of the process the stage where a person uses an aesthetic description about a work he is contemplating. As in the case of emotions, the naming or tagging of the preceding process when we need to clarify the episode for ourselves or someone else, we usually do this by conceptual and linguistic means. It is then when we look for the best candidate to communicate a summary of this process. Aesthetic phenomena can be considered as a process in which its stages and parts are constituted by modes of attention, perception, and second level, (abstract) integration perception into of forms, as well as their associate emotional episodes, and an output stage where the previous stages are articulated first conceptually and then linguistically. In this way, the final part of an aesthetic process is a sub-process where we constrain the selection of term attending to the characteristics of the aesthetic process. In this same stage a system of logic-linguistic (term-logic approach) constraints are used to further refine the restrictions on the selection of terms. This sub-process of attaching a tag to an aesthetic episode is terminal in the sense that it cannot lead to anything else except going into the aesthetic-process again, in this way we can integrate both the logic of criteria looks-is and the characteristic of being terminal proposed by Kivy.

My proposal is thus, that an aesthetic response can be seen as a process that develops through the following stages:

1) Sensory input
2) Recognition of the stimulus as an artefact (something intended, something man-made with a purpose)
3) Deliberated Attention Focusing
4) Form Perception and Integration
5) Basic Emotive Response
6) Cognitive Emotive Response
7) Second Order Form Integration, SOFI (Recognition of the abstract properties)
8) Second Order Cognitive Emotive Response (SOFI Associate emotive response)
9) Personal, Conceptual, and Linguistic Articulation of the Aesthetic-Episode

This process can be sometimes followed by a "feedback" stage:

10) Output Feedback, Assimilation of Personal to Culture

Stage 9 deals directly with our topic here. Tagging or naming the process itself. It consist in trying to differentiate, to make it specific in a way that it fits as recognizable, as a related to past experiences with art, or at least to fit it in our conceptual framework about what is art. That is, in this stage the person going trough the aesthetic episode tries to establish a link between the perceived object as it is represented *inside* the process (it does not deal with the stimulus as it is presented to the senses, but rather with a second order representation) and a property. This link must be such that can be recognized by anyone with a cultural framework similar to the person who experienced the aesthetic episode. An important fact is that the representation involved in this sub-process is abstract in nature; this fact is a necessary step if we are endorsing a notion of forms as described above. This internal formalist representation is linked to a tag in order to make sense of the internal process either for oneself or for someone else. A tag attached to the experience enable us to make a quick classification of the experience into a familiar context (our past experiences and concepts of what is aesthetically valuable for example) and at the same time it enable us to share it with other person by means of language.

Aesthetic Terms as *Locally* Terminal Tags

Descriptions of the type "O is A", where O refers to the object of perception (a painting, a symphony, a sculpture, etc) and A is an appropriate aesthetic term, have the function of articulating the perceived fitting-in of the represented object as part of a cultural category of stimuli; aesthetic stimuli. By using the expression "O is A" the speaker make public his perceived articulation. However he is not trying to just carry the information contained in "O is A" from one person to another, but rather to make his interlocutor aware of his internal articulation. In doing this, metaphors can fare better than literal descriptions, that is the reason why non-aesthetic terms are often used in this kind of aesthetic descriptions. Metaphors or other figures are used in order to make an efficient and clarifying summary of our internal perception to our interlocutor.

We can now summarize some features of the terms involved in the description resulting from this tagging sub-process. First, the stage of naming or tagging an aesthetic experience consists in trying to find a way to make oneself or someone else aware of the general result of a personal internal process. Second, there is no set of "characteristically" aesthetic terms, but they are constrained in two ways: by the particular characteristic of the specific aesthetic-process in which they are involved, and by the logic and linguistic characteristics inherent in sentences that communicate internal subjective impressions. Third, the aesthetic perception process includes cognitive task that are usually based on and modulated by the knowledge and experiences of the person going through the aesthetic episode; in this way culture plays a crucial role. Fourth, aesthetic descriptions have the only function of making ourselves or someone aware of our internal process; this is the reason why they are terminal in the sense of not leading to a further logical conclusion or a future action. Fifth, once the tag for a process has been attached, the description that conveys that attachment becomes part of the personal and collective aesthetic experience (in the sense of an episode from which we can lead); it becomes part of the culture that will play a role in future aesthetic-episode processes of other persons.

This last aspect calls for a revision of the idea of aesthetic terms as terminal terms. I propose that aesthetic term must see not as simple terminal but as *locally* terminal.

I will explain what I mean by locally terminal: In my model of aesthetic process aesthetic terms can lead to further aesthetic episodes. For example, if the aesthetic description is an art-criticism description which is meant to explain a work of art that would be otherwise not understandable. This critical description would make the work available to its audience, and, then, this critical description led to further aesthetic processes. In this case aesthetic descriptions get engaged in further aesthetic processes. This seems introduce a tension with the idea of term being characteristically terminal. I believe there is just a partial tension. They are terminal in the sense of being the output, the final stage of an internal process intended only to convey awareness, in this sense they do not lead to further conclusion or actions. But they become part of our cultural "aesthetic stock", so they necessarily participate in further aesthetic process. An aesthetic term is in this fashion *locally* terminal; it is the final stage of a private process, but it is also non-locally *fertile*, let us remember that the model proposed above contemplates a stage 10, where personal description gets assimilated in the culture: It is an essential characteristic of aesthetic terms to become part of our culture and get engaged in further aesthetic processes.

To summarize, an aesthetic term is a tag that is attached to an aesthetic process during its final, output stage. As a result aesthetic terms are locally terminal and potentially fertile.

Conclusion

The relevance of this notion of aesthetic term is that it allows us to account for two facts regarding a project of a face-value aesthetics in mathematics.

First, as Rota recognized, there are certainly epistemic phenomena involved in the reasons that lead mathematicians to make aesthetic descriptions. But we need not to jump to the conclusion that they are deceiving some epistemic quality. Rather we can see aesthetic phenomena as processes involving perception, emotion, cognition and culture. Among Cognition and culture mathematical knowledge-related issues play a central role.

Second, aesthetic phenomena in mathematics are often part of a larger chain of arguments and actions. For example, when a

mathematician qualifies a mathematical proof as an ugly proof, this often results in other mathematicians devising alternative less ugly proofs. There is a sense in which aesthetic descriptions are calls for further action to correct ugliness in mathematics. This contradicts the original terminal approach. However, in my locally terminal approach aesthetic descriptions can be seen as leading to further mathematical aesthetic episodes (trying to avoid ugliness is certainly a chain of actions but they are guided by internal aesthetic episodes). Leading to further aesthetic episodes is consistent with the idea of aesthetic terms being locally terminal and non-locally fertile.

Third, as McAllister has pointed out, past aesthetic descriptions have an impact in future ones. McAllister's aesthetic induction can have a place in the concept of aesthetic term as locally terminal presented here. Aesthetic induction can be seen as one of the mechanisms by which personal aesthetic description and terms are integrated in culture and how they are able to get engaged in further aesthetic process.

My aim here was to discuss the possibility of presenting a literal, face value, interpretation of aesthetic descriptions in mathematics. I have presented a first sketch of how this can be done and even how it can integrate some advantages of non-literal views. I believe I have established that proposals like Rota's or McAllister's are not face value approaches to aesthetic descriptions in mathematics. I also believe that I have showed that it is possible to think of a face value approach that complies with the most relevant characteristics attributed to aesthetic terms. Further development and research is necessary, but this is a research in progress and the first results presented here seem to be encouraging.

Bibliography

Ekman, P. (1994). "Moods Emotions and Traits" in Ekman, P.; Davidson, R. J. (eds), *The Nature of Emotion: Fundamental Questions*, New York: Oxford University Press.

_____. (2001). *Telling Lies: Clues to Deceit in the Marketplace, Politics, and Marriage*, New York: Norton.

_____, Davidson, R. J. (eds.) (1994). *The Nature of Emotion: Fundamental Questions*, New York: Oxford University Press.

_____, Scherer, K. R. (eds.) (1984). *Approaches to emotion*, Hillsdale, N.J : Erlbaum Associates.

Goodman, Nelson (1968). *Languages of Art: An Approach to a Theory of Symbols*. London: Oxford University Press.

Hanslick, Eduard (1986). *On the Musically Beautiful: A Contribution towards the Revision of the Aesthics of Music*. Hackett Pub. Co.

Hungerland, Isabel C. (1963). "The Logic of Aesthetic Concepts", *Proceedings and Addresses of the American Philosophical Association*, Vol. 36. (1962-1963), pp. 43-66.

Hyde, William H. (1978). "What Else Makes Aesthetic Terms Aesthetic?", Philosophy and Phenomenological Research, Vol. 39, No. 1, pp. 124-130.

Kivy, Peter (1975). "What Makes "Aesthetic" Terms Aesthetic?", *Philosophy and Phenomenological Research*, Vol. 36, No. 2, pp. 197-211.

_____. (2002). *Introduction to a philosophy of music*. Oxford: Clarendon Press.

Kuipers, Theo A. F. (2002). 'Beauty, a Road to The Truth' . Synthese 131 (3), 291-328.

LeDoux, Joseph E. (1994, 1). "Cognitive-Emotional Interaction in the Brain", in Ekman, Paul; Davidson, Richard J. (eds.) (1994), *The Nature of Emotion: Fundamental Questions*, New York: Oxford University Press.

_____. (1994, 2). "Emotion-Specific Physiological Activity", in Ekman, Paul; Davidson, Richard J. (eds.) (1994), *The Nature of Emotion: Fundamental Questions*, New York: Oxford University Press.

_____. (1996). *The Emotional Brain: the mysterious underpinnings of emotional life*, New York, N.Y.: Simon & Schuster.

McAllister, James (1996). *Beauty and Revolution in Science*. Ithaca, NY: Cornell University Press.

Meyer, Leonard B. (1968). *Emotion and Meaning in Music*. Chicago: The University of Chicago Press, 1968.

Mithen, Steven J. (2005). *The singing Neanderthals: the origins of music, language, mind, and body*. London: Weidenfeld & Nicolson.

Robinson, Jenefer (2005). *Deeper than Reason: Emotion and its Role in Literature, Music and Ar*t. Oxford University Press.

_____. (2004). "The Emotions in Art" in Kivy, Peter, ed., *The Blackwell Guide to Aesthetics*. Blackwell.

Rota, Gian-Carlo (1997). "The Phenomenology of Mathematical Beauty", *Synthese* 111: 171–182.

Scruton, Roger (1997). *The Aesthetics of Music*. Oxford: Clarendon Press.

INSTRUMENTALISMO VERSUS REALISMO
A CRISE NA FÍSICA DO SÉCULO XX

Rui Nobre Moreira
SAHFC/FCUL-CFCUL

Um físico de partículas, Oldershaw[1], num artigo publicado em 1988 afirmava que a física da segunda metade do século XX tinha-se entretido a construir teorias inverificáveis de 1ª ordem e teorias inverificáveis de 2ª ordem.

> **1ª ordem** – "Teorias" que só podem ser testadas quando se atingirem níveis de energia que não se podem atingir! Mesmo o actual LHC que alguns ainda têm a esperança de vir a evidenciar a existência do chamado bosão de Higgs, não vai explicar-nos, como é por demais evidente para os que conhecem as antinomias de Kant, a forma como o Universo foi criado. Seria necessário, partindo de uma visão correcta do que é a ciência, informar o comum cidadão que o chamado "modelo standard" é um, cada vez mais, longínquo projecto de teoria que, por exemplo, possui 19 parâmetros livres! Seria necessário que referissem todos os postulados explícitos e implícitos que estão na base das suas construções teóricas.
> **2ª ordem** – "Teorias" que dependem de tantos parâmetros (19 como acabei de referir no "modelo standard") que, qualquer que seja o resultado de uma experiência, podemos sempre acertar esses parâmetros para que a "teoria" passe a estar de acordo com esses resultados! Isto tem claras ressonâncias com modelos como o das esferas homocêntricas ou dos deferentes, excêntricas e epiciclos!

Alexandre Koyré afirmou que os cientistas investigam alegremente até que uma crise os obriga, finalmente, a meditar sobre os fundamentos daquilo que fazem. Hoje é por demais evidente que a física está em crise. Por isso é necessário regressar a "terreno firme"

[1] Robert L. Oldershaw, *American Journal of Physics*, Vol. **56** (12), December 1988.

para que possamos reconstruir um "edifício" que está em vias de ruir. Somos obrigados a repensar os fundamentos de tudo o que se tem feito em física ao longo do século XX.

É por tudo isto que aqueles que estudam os fundamentos da Física são obrigados a analisar mais de perto uma definição de teoria física. Mas uma definição de teoria física que nos facilite distinguir as teorias físicas consolidadas das que não passam de projectos de teoria. Para que isso se torne possível, é indispensável conhecer profundamente o estado actual da física, bem como a sua história.

Vamos escolher uma definição de teoria física que não é totalmente original mas que pode servir. Proponho seis critérios, que se podem subdividir em dois conjuntos: os três primeiros são critérios teóricos e os três últimos são critérios empíricos. Os critérios teóricos a que uma teoria deve obedecer são:

> **1 –** Partir necessariamente de um compromisso ontológico, isto é, admite que, à escala a que estamos a estudar os fenómenos, a realidade subjacente assume determinadas características. Este compromisso ontológico é sempre provisório mas é necessário, como tentarei mostrar ao longo deste trabalho. É sempre provisório porque entre o mundo e o que nós pensamos sobre ele nunca existe uma relação de identidade, mas apenas uma relação de analogia.
> **2 –** Desse compromisso ontológico emergem os postulados, ou seja, os postulados têm de ser coerentes com esse compromisso ontológico.
> **3 –** Esses postulados servem de suporte a toda uma estrutura matemática que permita quantificar a descrição dos fenómenos pretendidos.

Os três critérios empíricos são:

> **4 –** Tem de existir concordância quantitativa com os fenómenos que a teoria física pretende descrever. Esta concordância nunca é absoluta porque a uma medida está sempre associado um determinado erro, mesmo que esse erro possa ser, por vezes, muito pequeno. Mas entre ser pequeno e ser nulo a diferença é abissal. Este constitui apenas o primeiro critério empírico de validação de uma teoria física.

5 – Tem de possuir a capacidade de previsão de fenómenos até aí insuspeitos. Exemplos:

a) Previsão da existência do planeta Neptuno (mecânica newtoniana)

b) Ponto luminoso de Poisson (óptica ondulatória)

c) Efeito de túnel para partículas materiais (mecânica quântica)

Este é o segundo critério empírico de validação de uma teoria física.

6 – Tem de se mostrar apta a aumentar a nossa capacidade de acção no mundo, através da construção de instrumentos até aí impossíveis de conceber. Uma teoria não é apenas um instrumento que nos permite descrever o que observamos como defendem os instrumentalistas. Uma teoria é, se me permitem a metáfora, mais uma máquina ferramenta que nos permite construir novas ferramentas, ou instrumentos, com que alargamos a nossa capacidade de acção no mundo. Atrever-me-ia a dizer que podem existir instrumentos sem teoria, mas que não podem existir teorias sem este tipo de instrumentos. Este é o terceiro critério empírico de validação de uma teoria física. Isto significa que, de acordo com este critério empírico, em física, nenhum projecto de teoria que não tenha ultrapassado esta derradeira fronteira deve ser considerado uma teoria física.

Vamos considerar vários exemplos de teorias e projectos de teorias (antigos e actuais), e verificar a quantos destes critérios obedecem.

Modelo matemático dos deferentes, excêntricas e epiciclos

1 – Sim
2 – Sim
3 – Sim
4 – Sim
5 – Não
6 – Não

Mecânica e gravitação newtonianas

1 – Sim
2 – Sim
3 – Sim
4 – Sim

5 – Sim
6 – Sim

Termodinâmica

1 – Sim
2 – Sim
3 – Sim
4 – Sim
5 – Sim
6 – Sim[2]

Electromagnetismo maxweliano

1 – Sim
2 – Sim
3 – Sim
4 – Sim
5 – Sim
6 – Sim

Mecânica quântica

1 – Sim
2 – Sim
3 – Sim
4 – Sim
5 – Sim
6 – Sim[3]

Relatividade restrita

1 – Sim
2 – Sim
3 – Sim
4 – Sim
5 – Sim
6 – Não[4]

[2] Frigoríficos, ar condicionado, bombas de calor.

[3] Díodos, transístores, computadores de pequenas dimensões mas com memória e velocidade de cálculo enormes, microscópicos de efeito de túnel, etc.

[4] Pode-se resolver o problema do acerto dos relógios do sistema GPS sem recorrer à relatividade restrita (Franco Selleri, 2003). Além disso, foi Henri Poincaré quem primeiro escreveu a célebre equação $E = m\,c^2$. Não precisou da teoria da relatividade restrita para tal. Sabemos hoje que existem outras transformações em

Relatividade geral

1 – Sim
2 – Sim
3 – Sim
4 – Sim
5 – Não
6 – Não

Electrodinâmica quântica

1 – Sim
2 – Sim
3 – Sim
4 – Sim
5 – Não
6 – Não

Modelo "standard"

1 – Sim
2 – Sim
3 – Sim
4 – Sim
5 – Não
6 – Não

Podemos concluir, por este quadro, que a mecânica quântica foi a última teoria que conseguiu respeitar todos estes critérios. Qual a razão porque, depois dela, nunca mais os físicos conseguiram construir teorias que tivessem aumentado a nossa capacidade de acção no mundo através da construção de novos instrumentos? Vejamos o problema um pouco mais de perto.

Quer a mecânica quântica, quer a relatividade restrita, pretenderam, cada uma à sua maneira, criar uma teoria que "conciliasse" a mecânica newtoniana e o electromagnetismo maxwelliano, as duas grandes teorias da física anterior ao século XX.

Quanto à relatividade restrita: A relatividade restrita optou por modificar a mecânica newtoniana de tal forma que a nova mecânica fosse invariante para as transformações de Lorentz. Einstein

---

que a velocidade da luz não é a mesma em todos os referenciais e para as quais a equação $E = m\,c^2$ continua válida. (Franco Selleri, 2003)

justificou-o com o argumento de que o electromagnetismo maxwelliano descrevia fenómenos mais precisos que a mecânica newtoniana. Foi essa a razão que o levou a optar por alterar a mecânica e não o electromagnetismo. Os dois postulados que suportam a relatividade restrita são:

> **1 -** Todas as leis físicas são as mesmas em todos os referenciais de inércia, isto é, quer a mecânica, quer o electromagnetismo, devem manter a sua capacidade de descrever os fenómenos observados em qualquer referencial de inércia e, por essa razão, as equações que as constituem devem manter a mesma forma em todos esses referenciais. Para isso, as equações que usamos nessas teorias devem obedecer às mesmas leis de transformação quando passamos de um referencial para outro. Foi por esta razão que, tal como disse no início deste parágrafo, Einstein alterou a mecânica newtoniana para que esta se tornasse invariante para as transformações de Lorentz, já que estas transformações eram as que mantinham invariantes as equações de Maxwell.
> **2 -** O outro postulado da relatividade restrita diz-nos que a velocidade da luz é a mesma em todos os referenciais de inércia. Como corolário, a relatividade restrita conclui que a velocidade da luz é a velocidade máxima com que podemos transmitir informação.

Problemas que emergiram recentemente:

> **1 -** No livro que escreveu em 2003[5], Franco Selleri chama a atenção para que podemos construir transformações diferentes das de Lorentz que não impõem que a velocidade da luz seja a mesma em todos os referenciais. Na construção da teoria da relatividade restrita Einstein admitiu que a velocidade de ida e volta da luz medida num referencial em movimento era a mesma nos dois sentidos. A conhecida "maldição" de Poincaré afirmava a impossibilidade de medir a velocidade da luz num só sentido. Como acontece frequentemente, afirmações tão peremptórias

[5] Franco Selleri, *Lezioni di relatività da Einstein all'etere di Lorentz,* Progedit, Bari, 2003.

arriscam-se a ser refutadas. Foi isso que aconteceu quando José Croca, num artigo publicado em conjunto com Franco Selleri[6], mostrou que tal essa pretensa impossibilidade não passa de ilusória. Somos obrigados a concluir que a hipótese admitida por Einstein passa a poder ser testada empiricamente. Pode ser que não se confirme… Convém acrescentar que, para as transformações propostas por Selleri, a célebre equação $E=mc^2$ continua válida.

2 - Hoje é já evidente que se conseguiu transmitir informação a velocidades superiores à velocidade da luz! A 40ª Sinfonia de Mozart foi transmitida à velocidade de 4.7 c![7] Isto foi conseguido em condições de efeito de túnel.

Tudo isto leva-nos a concluir que temos de ser muito cuidadosos quando analisamos um projecto de teoria que ainda não conseguiu satisfazer todos os 6 critérios que referi antes. Não quero com isto dizer que a relatividade restrita não forneceu ganhos cognitivos importantes. Estou apenas a dizer que a crise obriga-nos a olhar para tudo isto com muito mais cuidado.

Quanto à mecânica quântica: A mecânica quântica também pretendeu "conciliar" a mecânica de Newton e o electromagnetismo de Maxwell. Mas desta vez pretendeu "conciliar" a ontologia em que se fundava cada uma das duas teorias clássicas. Como veremos essa "conciliação" levará a uma posição de negação de qualquer uma dessas ontologias. Por um lado, a mecânica de Newton descrevia o movimento de corpúsculos num espaço eminentemente vazio, por outro, o electromagnetismo de Maxwell descrevia ondas que se propagavam num espaço inevitavelmente cheio para suportar essas mesmas ondas. Na primeira existiam interacções à distância com velocidade infinita, na segunda existiam apenas interacções de ponto para ponto contíguo à velocidade da luz. Como foi isto feito?

Esta "conciliação" iniciou-se através de dois percursos heurísticos diferentes: Na universidade de Göttingen, Max Born, Heisenberg. Jordan, entre outros, pretenderam criar uma mecânica

[6] José Croca e Franco Selleri, *Is the one-way velocity of light measurable?* Nuovo Cimento, 114 B, pp. 447-457, 1999.

[7] Günter Nimitz, *Superluminal Signal Velocity and Causality,* Foundations of Physics, Vol. 34, No. 12, December 2004.

formalmente decalcada da mecânica de Newton. Como o fizeram? Utilizaram as equações da mecânica de Newton e substituíram as grandezas clássicas por matrizes. Daí recebeu o nome de mecânica das matrizes. Para a mecânica das matrizes existiriam apenas corpúsculos. Mas corpúsculos aos quais, contrariamente ao que acontecia na mecânica de Newton, estaria associada uma incapacidade de determinar simultaneamente uma posição e uma velocidade bem precisas. Foi neste contexto que Heisenberg chegou às suas célebres relações de indeterminação.

Satisfeito com esse resultado, foi a Copenhaga mostrá-lo a Niels Bohr em Fevereiro de 1927. Este respondeu-lhe que o que ele tinha feito não passava de um truque matemático e que tudo aquilo tinha de ser interpretado em termos físicos! Heisenberg saiu de lá um pouco combalido. Niels Bohr já era Prémio Nobel (por causa do seu modelo atómico de 1913) e isto constituiu um duro golpe no jovem Heisenberg.

Porque aconteceu isto? Porque Bohr já sabia que o vienense Schrödinger, então a trabalhar na Suíça, tinha seguido outro percurso heurístico. Schrödinger tinha utilizado algo já detectado por Hamilton no século XIX. Este tinha chamado a atenção para o facto de existir uma enorme similitude entre a equação de Hamilton-Jacobi, um formalismo alternativo ao de Newton para a sua mecânica, e a equação da óptica geométrica, a chamada equação do eikonal. Esta equação é uma simplificação da equação de onda para a propagação das ondas electromagnéticas quando o comprimento de onda tende para zero (frequência tende para infinito).

Essa equação é:

$$\left(\frac{\partial S}{\partial x}\right)^2+\left(\frac{\partial S}{\partial y}\right)^2+\left(\frac{\partial S}{\partial z}\right)^2=n^2(x,y,z) \qquad (1)$$

Em que *S* representa as frentes de onda e, na aproximação da óptica geométrica, os raios de luz são perpendiculares a estas. Podia-se pensar que, nesta aproximação, estaríamos a representar os corpúsculos de luz da óptica corpuscular de Newton que a esta escala funciona.

A equação de Hamilton-Jacobi tem a forma

$$\left(\frac{\partial S}{\partial x}\right)^2+\left(\frac{\partial S}{\partial y}\right)^2+\left(\frac{\partial S}{\partial z}\right)^2+V(x,y,z)=-\frac{\partial S(x,y,z)}{\partial t} \qquad (2)$$

Esta equação descreve trajectórias de corpúsculos que obedecem à mecânica de Newton as quais, tal como na equação anterior, são perpendiculares à superfície *S* solução desta equação.

Mas, como disse, a equação do eikonal é uma simplificação da equação de onda do electromagnetismo que, como se sabe, é:

$$\frac{\partial^2\psi}{\partial x^2}+\frac{\partial^2\psi}{\partial y^2}+\frac{\partial^2\psi}{\partial z^2}=\frac{1}{c^2}\frac{\partial^2\psi}{\partial t^2} \qquad (3)$$

Simplificação que é possível quando, tal como referi acima, os comprimentos de onda tendem para zero. O problema que Schrödinger queria resolver era o de encontrar uma equação de onda que estivesse para a equação de Hamilton-Jacobi, tal como a equação de onda do electromagnetismo estava para a equação do eikonal.

Foi este o percurso heurístico que Schrödinger seguiu para chegar à sua célebre equação. Chamo a atenção que não se chega à equação de Schrödinger através de uma demonstração. A equação de Schrödinger, que a seguir escrevemos, constitui um postulado da mecânica quântica. Os postulados, por definição, não se demonstram.

$$-\frac{\hbar^2}{2m}\left(\frac{\partial^2\psi}{\partial x^2}+\frac{\partial^2\psi}{\partial y^2}+\frac{\partial^2\psi}{\partial z^2}\right)+V(x,y,z)\,\psi=i\hbar\frac{\partial\psi(x,y,z)}{\partial t} \qquad (4)$$

Schrödinger estava, assim, à procura de uma mecânica ondulatória em que, como o seu próprio nome indica, só existiriam ondas!

Niels Bohr sabia disto. Sabia também que Schrödinger, numa série de artigos publicados em 1926, mostrara que o seu formalismo era equivalente ao que tinha sido desenvolvido em Göttingen! Foi por saber tudo isto que Niels Bohr foi obrigado a pensar na forma de integrar num determinado quadro conceptual as duas concepções que

tinham surgido por duas vias heurísticas aparentemente independentes.

Foi no período entre Fevereiro de 1927 (visita de Heisenberg) e o Verão desse mesmo ano que Niels Bohr chegou à sua interpretação, dita ortodoxa, de Copenhaga, ou bohriana do formalismo quântico. Sabia que nas equações fundamentais da física quântica,

$$E = \hbar\omega \qquad \textit{Max Planck}$$

$$\vec{p} = \hbar\vec{k} \qquad \textit{Louis de Broglie}$$

as características ondulatórias e as características corpusculares estavam indissoluvelmente ligadas. Foi, como disse, durante este período que Niels Bohr conseguiu transpor para a física o *princípio de complementaridade*. Um princípio que o seu professor de filosofia, amigo de seu pai e seu próprio amigo, Harald Høffding, já tinha introduzido em psicologia e em filosofia. Antes de o enunciar penso que é útil construir uma frase em que o início é de Harald Høffding e o final é de Niels Bohr:

> *"Existe um resíduo irracional [irredutível - R.M.] (Høffding)[8], que está agora expresso de uma forma matematicamente lúcida[9]" (Bohr)*

Høffding pensou que detectara uma relação de complementaridade entre as primordiais, segundo ele, funções psicológicas que são ver e compreender. Segundo Høffding, ver e compreender seriam duas *necessidades inconciliáveis* do espírito humano que este jamais conseguiria ultrapassar. Em defesa desta sua posição Høffding cita Poincaré:

> *"Continuidade e descontinuidade são correlativos que se suprem um ao outro. Designam diferentes pontos de vista e*

[8] H. Høffding; É. Meyerson, *Correspondence entre Harald Høffding et Émile Meyerson*, ed. Frithiof Brandt, Hans Høffding et Jean Adigard des Gautries, (Einar Munskgaard), Copenhaga, 1939, p. 153.

[9] Relatada por Heisenberg in, S. Rozental (ed.), *Niels Bohr. His life and work as seen by his friends and colleagues*, North-Holland Publishing Company – Amsterdam, pp. 106-107.

> *diferentes operações; a história das ciências mostra como, tanto uma, como a outra, tomam a primazia, mas de tal modo que a luta entre elas recomeça sempre. Ninguém deitou sobre a sua relação uma luz mais esclarecedora do que Henri Poincaré quando disse: "Esta luta durará enquanto fizermos ciência, enquanto a humanidade pensar, por que ela é devida a duas necessidades inconciliáveis do espírito humano, das quais esse espírito não se pode despojar sem deixar de existir, a de compreender e nós não podemos compreender senão o finito, e a de ver, e nós não podemos ver senão a extensão que é infinita..."* [10]

Høffding considerava a psicologia como uma fenomenologia do espírito e, por isso, todas as leis que se detectassem no funcionamento espontâneo do espírito humano, que é o objecto do estudo da psicologia, espraiar-se-iam inevitavelmente para as formas mais elaboradas de funcionamento desse mesmo espírito. Todas elas teriam de ser *psicologicamente possíveis*[11]. Segundo Høffding as categorias prestariam inevitavelmente vassalagem às funções psicológicas atrás referidas, e às relações que entre elas se verificavam, segundo ele, em psicologia. Daí a estabelecer uma relação de complementaridade entre as formas *a priori* da sensibilidade (espaço e tempo) e as formas *a priori* do entendimento de Kant (categorias – das quais é inevitável destacar a de causalidade) foi um pequenino passo. Høffding, como acabei de afirmar, transpôs essa relação de complementaridade da psicologia para a filosofia. Niels Bohr tratou de a transpor para a física atribuindo-lhe o nome de princípio de complementaridade. É por essa razão que encontramos logo na primeira exposição que Niels Bohr faz da sua interpretação do formalismo quântico, no Congresso Volta de 1927 (Lago de Como)[12] a afirmação de que nunca poderemos atingir

[10] H. Høffding, *La Relativité Philosophique* 1925, pp. 197-199. Esta tradução francesa engloba dois ensaios em dinamarquês de Høffding: *Totalitatet som* Kategori (Totalidade como categoria), Det kgl. D. Vid. S. Skrifter. 6. Række, hist. Filosofisk afd. 2. Copenhaga 1917, e *Relation som* Kategori (Relação como categoria), D. kgl. D. Vid. S. Filosofisk Meddelser, I, 3, Copenhaga, 1921. A frase de Poincaré que Høffding cita pode ser encomtrada em *"Les Conceptions Nouvelles de la matière"*, em *"Le matérialismo actuel"*, p. 67.

[11] H. Høffding, *Esquisse d'une psychology fondée sur l'expérience*, Alcan, Paris, 1903, p. 362.

[12] Congresso Internazionale dei Fisici, Lago di Como, 1927

simultaneamente uma descrição espácio-temporal dos fenómenos físicos (quânticos) e uma descrição causal (desses mesmos fenómenos).

Foi aqui que o maior problema epistemológico do século XX emergiu. Mas, sendo a mecânica quântica a melhor teoria que o homem conseguiu construir até hoje, obedecendo ela a todas as condições que atrás enunciei, podemos então perguntar: onde está o problema? O problema reside em que quando os físicos se começaram a confrontar com fenómenos que, quer a relatividade restrita, quer a mecânica quântica, pretendiam descrever, muitos acabaram por convencer-se de que era impossível descrever esses fenómenos no velho quadro do espaço e do tempo e, simultaneamente, agora apenas na mecânica quântica, que o conceito de causalidade deixaria de se poder aplicar com a generalidade com que se tinha feito na física até então. Foi neste quadro que os físicos acabaram por cair numa posição instrumentalista ou simbolista. Aqui, o termo instrumentalista assume o significado usual, ou seja, o de encarar uma teoria como um instrumento que permite fazer previsões empíricas e outros fins práticos, sem que nos comprometamos seriamente, mesmo que provisoriamente, com uma qualquer ontologia prévia. Mesmo o empirismo construtivo de um autor como Bas Van Fraasen mais não faz do que aceitar, eu atrever-me-ia a dizer displicentemente, uma teoria, sem nunca acreditar verdadeiramente nela. Quando digo acreditar não quero dizer acreditar que essa teoria seja uma fiel descrição da coisa em si, mas que essa entre a coisa em si e o que pensamos sobre ela, que é o que, na sua essência, essa teoria exprime, existe alguma sintonia, pelo menos à escala em que estamos a descrever os fenómenos. É exactamente neste ponto que a distinção entre uma posição realista e uma posição instrumentalista tem necessariamente de emergir. Apesar de uma aproximação que se tem observado entre essas duas posições, que tinham partido de atitudes radicais, aproximação essa que leva, por exemplo, a que um realista acabe por aceitar que entre o que ele pensa sobre o mundo e esse mesmo mundo existe apenas uma relação de analogia e nunca uma relação de identidade, há entre as duas posições em confronto uma divergência residual que tem consequências significativas como me esforçarei por mostrar ao longo deste trabalho. O terceiro critério empírico, que qualquer instrumentalista poderia aceitar, nomeadamente o próprio introdutor do termo, John Dewey, é o que

serve para indiciar a existência dessa relação de analogia entre o mundo e o que nós pensamos sobre ele e, além disso, definir uma heurística positiva, no sentido lakatiano do termo, muito geral, que acredito poder levar a um mais rápido entendimento do mundo que nos rodeia.

Como dizia, Bohr assumiu uma atitude instrumentalista ou simbolista. Poderia, a este propósito, citar uma afirmação de Harald Høffding numa sua carta para Émile Meyerson

> "…Numa conversa que manteve comigo depois da conferência, Bohr disse-me que está cada vez mais convencido da necessidade da simbolização se quisermos exprimir os últimos resultados da física."[13]

Simbolismo refere-se à escolha de símbolos que representam algo de abstracto e não pretendem que se lhes atribua um significado literal, quer dizer, relacionados com algo de real. Desta forma, os físicos defenderam, com Bohr na vanguarda desse movimento, que as teorias seriam meros instrumentos a que não podemos exigir mais que a descrição dos fenómenos, sem que lhes possamos atribuir nenhum conteúdo realista. Mas, para Bohr, a forma como apreendemos os fenómenos está conformada pelas características que detectamos no funcionamento espontâneo do espírito humano, objecto do estudo da psicologia. Psicologia onde o seu professor de filosofia pensava ter detectado uma relação de complementaridade entre as funções psicológicas ver e compreender que, inevitavelmente, passaria para a filosofia (Høffding já o tinha feito) e para a ciência. Volto a repetir que isto seria inevitável pois, como dizia Höffding, toda a actividade do espírito tem de ser *psicologicamente possível*. Os símbolos a que Bohr se referia, na frase anteriormente citada, relacionavam-se muito mais com a representação das nossas incapacidades cognitivas do que com algo de real subjacente aos fenómenos estudados. O anti-realismo bohriano, ou como defende J. Faye[14], o anti-realismo objectivo de Niels Bohr começa aqui.

---

[13] H. Høffding; É. Meyerson, *Correspondence entre Harald Høffding et Émile Meyerson*, ed. Frithiof Brandt, Hans Høffding et Jean Adigard des Gautries, (Einar Munskgaard), Copenhaga, 1939. Carta de 30 de Dezembro de 1926.

[14] J. Faye, *Niels Bohr: His Heritage and Legacy*, Kluwer Academic Publishers, Dordrecht, 1991.

O enorme sucesso da mecânica quântica enquanto última teoria digna desse nome que o homem construiu, e o prestígio de Bohr enquanto físico teórico, fizeram com que a posição deste prevalecesse, não só entre os epistemólogos que, à partida, concordavam com uma visão instrumentalista das teorias físicas, como entre outros que levam demasiado a sério aquilo que alguns cientistas lhes dizem. Esta é uma das razões porque muitos epistemólogos do século XX acabaram por aceitar como inevitável a adopção de uma visão instrumentalista das teorias físicas. Chegam ao ponto de defender que as posições instrumentalista e realista são equivalentes. Mas a física e as suas maiores construções, as teorias, respondem à necessidade premente e permanente de aumentar a probabilidade de sobrevivência da espécie que a cria. Aumentar a capacidade de sobrevivência da espécie humana passa, não apenas pela possibilidade de estabelecer nexos causais entre fenómenos e de os integrar numa descrição espácio-temporal que nos permitem prever acontecimentos futuros, por mais falível que essa previsão possa ser, como, acima de tudo, por permitir a construção de novos instrumentos que aumentem a nossa capacidade de agir no mundo. Instrumentos que seriam, na verdadeira acepção da palavra, inconcebíveis sem essas teorias. Porque o seu funcionamento depende de tudo o que define essa teoria, ou seja, do compromisso ontológico, provisório é certo, que está na sua base, e dos postulados compatíveis com esse mesmo compromisso. Ora a física da segunda metade do século XX, em que a posição instrumentalista entre os físicos, frequentemente a mais radical, foi dominante, nada mais fez que construir projectos de teoria que se têm revelado estéreis. Estéreis porque nos mantiveram incapazes de, com base neles, construir instrumentos que tivessem aumentado a nossa capacidade de agir no mundo. Na versão mais radical do instrumentalismo dos físicos na segunda metade do século XX perdeu-se a necessidade que o introdutor do termo, John Dewey, realçava na ciência que era a necessidade de resolver problemas práticos. Mas Dewey falava de ciência e não de teoria. Esta precisão é importante porque, focando a nossa atenção nas teorias, que podem integrar ou não um programa de investigação científica, e não na ciência, nomeadamente na física, ganhamos mais um tipo de critério empírico para distinguir entre um projecto teoria e uma teoria. As teorias físicas, as que obedecem aos seis critérios enunciados no início deste texto, continuam a respeitar este critério, mas isto não pode servir de camuflagem para que os

projectos de teorias não o respeitam tentem passar desapercebidos. É, por isso, necessário sublinhar, para evitar a confusão associada à propaganda que os defensores destes projectos de teorias lançam para o ar, quando falam do enorme desenvolvimento tecnológico entretanto conseguido, que os novos instrumentos que se foram construindo ao longo da segunda metade do século XX e início do século XXI, apoiam-se todos nas sólidas teorias que se tinham construído antes desse período, mecânica quântica incluída, e nunca nos projectos de teoria que se desenvolveram ao longo desse mesmo período.

Como disse, a última grande teoria que o homem construiu foi a mecânica quântica ortodoxa. Por isso, é importante saber que somos, hoje, capazes de explicar a razão porque a mecânica quântica ortodoxa conseguiu ter o enorme sucesso que inegavelmente teve, apesar de, na sua origem, ter partido de uma posição instrumentalista radical. A mecânica quântica é uma teoria linear que consegue descrever com o rigor necessário o comportamento estatístico de um grande conjunto de partículas. Foi aqui que a mecânica quântica obteve todo o seu inegável sucesso. Mas, como qualquer outra teoria, a mecânica quântica ortodoxa não é A TEORIA. Para um realista não ingénuo, a relação entre o mundo e o que nós pensamos sobre ele nunca é uma relação de identidade. É somente uma relação de analogia. Um realista não ingénuo sabe que o compromisso ontológico é sempre provisório mas, apesar disso, considera-o indispensável. Porque só esse compromisso nos permite relacionar as novas teorias com as *boas*[15] teorias que até então foram criadas. Nenhuma delas é A TEORIA. Nem elas, nem as que se lhes seguirem o serão. A realidade é seguramente mais complexa do que o que pensamos sobre ela. Mas o último critério que enunciei para a validação de uma teoria física é o que nos pode, de alguma forma, garantir alguma sintonia entre o mundo e o que nós pensamos sobre ele, ou seja, as nossas teorias, à escala dos fenómenos que elas pretendem descrever. Um físico, ao adoptar uma posição realista, deve olhar para toda a informação empírica disponível e para, e só para, as teorias físicas consolidadas, aquelas que nos possibilitaram construir novos instrumentos até aí inconcebíveis, e meditar no significado profundo que nelas está contido. Só um realista pode pensar desta maneira. Um realista não pode aceitar a incomensurabilidade entre teorias como defendeu

[15] As teorias que obedecem aos 6 critérios referidos no início deste texto.

Kuhn. Mesmo admitindo que essa incomensurabilidade seja apenas epistemológica e não ontológica em Kuhn, invocando que este defende, na sua obra *A estrutura das revoluções científicas*, que não é o mundo que se altera numa revolução deste tipo, mas apenas a visão do mundo dos cientistas, frase com a qual estou de acordo, a sua conjugação com uma visão instrumentalista radical de uma teoria física, que a sua visão da história da ciência acaba por promover, acaba por levar a um resultado, quanto a mim, nefasto. Dessa forma, seguir o fio condutor que tenho estado a defender, é a via mais promissora de estabelecer pontes entre diferentes teorias físicas. As *boas* teorias físicas são sempre comensuráveis. Pelo menos num sentido, ou seja, a antiga *boa* teoria está toda contida na nova *boa* teoria. Todos os fenómenos significantes para a antiga continuam a sê-lo para a nova. Algo que já Lakatos defendeu sem o cuidado de introduzir explicitamente o último critério de validação de uma teoria física que referi no início deste texto. Não existe, assim, a necessidade de, por exemplo, a nova "matriz disciplinar" proceder à limpeza que Kuhn diz acontecer sempre no início de um período de ciência "normal", logo após uma revolução científica. É, portanto, necessário distinguir vários tipos de revoluções em física. As que tratam da passagem de uma *boa* teoria física para outra *boa* teoria física, e as outras. Nestas últimas podemos detectar essa limpeza. Nas primeiras não.

A posição realista é incompatível com a aceitação de que a interpretação bohriana do formalismo quântico trace um limite definitivo à nossa capacidade de entender o mundo. Aceitar o princípio de complementaridade, é aceitar que este, de uma vez por todas, teria estabelecido que os conceitos de onda e corpúsculo são meras *necessidades inconciliáveis* do espírito humano, negando qualquer relação entre estes conceitos e uma realidade subjacente. A interpretação bohriana do formalismo quântico assume uma posição inevitavelmente ingénua. Pretender que uma teoria defina, de uma vez por todas, os limites da nossa capacidade de entender o mundo, só pode ser comparada com a convicção de que as nossas teorias são um espelho fiel de uma realidade subjacente, ou seja, que entre as nossas teorias e o mundo que elas pretendem descrever existe uma relação de identidade. Um realista não pode cair nesta ingenuidade, mas tem, inevitavelmente, de procurar o significado um pouco mais profundo,

nunca definitivo, que está contido no maior problema herdado da física do século XX, o *dualismo onda-corpúsculo*.

O físico que, no século XX, assumiu uma posição realista mais consistente foi Louis de Broglie. Para este físico francês, as partículas quânticas seriam simultaneamente constituídas por uma singularidade que manifestaria as propriedades corpusculares dessa partícula, e por uma onda que levaria essa partícula a manifestar propriedades ondulatórias. Assim, durante a segunda metade do século XX definiram-se claramente dois programas de investigação em física fundamental:

> - o programa de investigação científica de Niels Bohr e dos seus continuadores, que tem usufruído de enormes investimentos, quer em vidas de pessoas inteligentes, quer em meios financeiros (veja-se o custo do actual LHC da ordem de $8x10^9$ €!). Contudo, não conseguiram aumentar em nada a nossa capacidade de acção no mundo com os projectos de teorias que têm desenvolvido. Os sucessos que apresentam relacionam-se apenas com previsões da existência de partículas que quase sempre têm imposto uma revisão da teoria previamente existente. Ou seja, são teorias largamente construídas a posteriori como denuncia o artigo de Oldershaw citado no início deste trabalho. Acertar na lotaria depois do sorteio é fácil. O difícil é consegui-lo antes do sorteio. São, por isso, projectos de teoria pouco credíveis.
> - o programa de investigação científica de Louis de Broglie e dos seus continuadores que, com meios incomensuravelmente menores, conseguiu propor a realização de experiências que pretendem evidenciar que, por exemplo, o fotão tem estrutura. [16] E se há uma partícula elementar digna desse nome, essa partícula será indiscutivelmente o fotão.[17] Claro que a estrutura do fotão a que me refiro tem a ver com o simultâneo carácter corpuscular e ondulatório e não com a possibilidade de partir o corpúsculo que faz parte do fotão. Só para um realista integrado

[16] No caso do programa bohriano, mesmo com o LHC, o electrão continuará a ser uma partícula sem estrutura.

[17] Claro que a estrutura do fotão a que me refiro tem a ver com o simultâneo carácter corpuscular e ondulatório e não com a possibilidade de partir o corpúsculo que faz parte do fotão.

> no programa de investigação de Louis de Broglie é possível conceber experiências que procurem evidenciar essa estrutura. O que acabei de dizer relaciona-se com o que Imre Lakatos chamou a heurística positiva de um programa de investigação científica. É neste ponto que vemos claramente as diferentes consequências que existem entre adoptar uma posição instrumentalista e uma posição realista. Para esta última, ao aceitar que as ondas e os corpúsculos existem como algo de real, existe a necessidade de procurar entender mais profundamente o significado do *dualismo onda-corpúsculo*.

Podemos adiantar que, dentro do programa de Louis de Broglie, já foi evidenciado[18] que as relações de Heisenberg foram ultrapassadas na prática laboratorial. Isto representa um fortíssimo golpe na convicção de que o princípio da complementaridade, de que as relações de Bohr-Heisenberg são uma consequência, e que está na base de toda a interpretação bohriana do formalismo quântico, constitui um *resíduo irracional* que no formalismo quântico ortodoxo se exprime de uma *forma matematicamente "lúcida"*. Na base de todo o programa de investigação proposto por Bohr, está a aceitação da existência de um limite para o nosso entendimento do mundo microfísico. Alguns têm consciência disto. No terceiro volume das *Lectures on Physics* de Richard Feynman podemos ler

> *O princípio de incerteza "protege"a mecânica quântica. Heisenberg reconheceu que se fosse possível medir o momento e a posição simultaneamente com maior precisão, a mecânica quântica ruiria.*[19]

Não posso deixar de chamar a atenção para dois erros que esta frase encerra. Em primeiro lugar, o de Heisenberg quando confunde medição com previsão de uma futura medição. Já se sabe, há muito, que é possível medir a posição e o momento de uma partícula quântica com uma precisão superior à que as relações de indeterminação de

[18] J. Croca, *Experimental Violation of Heisenberg's Uncertainty Relations*, talk at the 5th UK Conference on the *Conceptual and Philosophical Problems in Physics*, at Oxford, Sept. 1996.
[19] Vol III, Cap. 1, p. 11.

Bohr-Heisenberg impõem[20]. Mas não é desta forma que elas devem ser compreendidas. Elas impõem que não se consegue *prever o resultado de uma futura medição* sobre a posição e o momento de uma partícula com uma precisão superior à que elas impõem[21]. Em segundo lugar, a mecânica quântica não ruiria, apenas se estabeleceriam os limites da sua aplicabilidade. No programa de investigação proposto por Louis de Broglie já foram deduzidas relações de incerteza mais gerais que as de Heisenberg, as quais prevêem a possibilidade de estas serem ultrapassadas. Voltarei a este tema mais adiante.

Por tudo o que acabei de dizer, não é lícito defender a equivalência entre uma posição instrumentalista, que considera uma teoria como um mero instrumento, que outra coisa não pretende que descrever os resultados das observações, sem admitir que exista uma relação entre essa teoria e o mundo por detrás dos fenómenos que observamos, e uma posição realista, que admite que existe alguma correspondência entre o mundo e o que nós pensamos sobre ele. Uma relação, como disse, de analogia e nunca uma relação de identidade.

Foi a posição instrumentalista defendida pela maioria dos físicos que se confrontavam com novos fenómenos no início do século XX, que levou muitos epistemólogos a defenderem, também eles, uma posição instrumentalista. Foi a posição instrumentalista que muitos físicos adoptaram, que fez com que se tivessem entretido a construir projectos de teoria de tipo epicíclico durante toda a segunda metade do século XX, processo que, infelizmente, perdura até hoje. Projectos de teoria que nunca conseguiram ultrapassar o último critério de validação de uma teoria física que apresentei no início deste trabalho. Não ultrapassaram nem se entrevê que consigam.

Paul M. Dirac disse, já numa idade avançada, que nos devíamos ter esquecido de alguma coisa ao longo do caminho. Isto foi dito pelo homem que deu a última contribuição para o fecho da abóbada do formalismo quântico ortodoxo: a equação de Dirac. Uma equação em

[20] O próprio Heisenberg vem, posteriormente, reconhecer isto.

[21] - Andrade e Silva, *Peut-on connaître simultanément et avec précision la position et la quantité de mouvement d'une particule?*, Portugaliæ Physica, 4 (1966) 257;
- K. Popper, *Quantum Theory and the Schism in Physics*, Hutchinson, London, 1972;
- J. Croca, *Towards a Nonlinear Quantum Physics*, World Scientific, 2003, pp. 34-37.

que tentou conjugar a mecânica quântica e a relatividade restrita. Mas foi uma conjugação mais formal do que real. Foi uma espécie de "pastiche" formal entre a relatividade restrita e a mecânica quântica. Penso que aquilo de que os físicos se esqueceram, encetando uma fuga para a frente, foi do grande problema que a física do século XX não resolveu sem abandonar uma posição realista. Estou a referir-me, como é evidente, ao problema do *dualismo onda-corpúsculo*. A física do século XX satisfez-se com os marabilismos epistemológicos de Niels Bohr, ao introduzir o seu princípio de complementaridade como pedra de toque de toda a interpretação do formalismo quântico ortodoxo. Como disse atrás, os conceitos de onda e corpúsculo seriam, para Bohr, meras necessidades psicológicas do homem e não corresponderiam a nenhuma realidade subjacente.

Hoje percebemos porque tudo isto aconteceu[22]. Tudo isto aconteceu porque todas as equações da mecânica quântica são equações lineares. Desde a equação de Schrödinger, passando pela de Pauli e pela de Klein-Gordon, acabando na equação de Dirac, todas elas são equações lineares. Como sabemos, estas equações constituem um dos postulados da mecânica quântica. Um postulado que, consoante o sistema quântico que queremos descrever, nos diz qual a equação que devemos utilizar. O que significa uma equação ser linear? Significa que ela descreve sistemas em que o todo é igual à soma das partes. Isto significa que a mecânica quântica, que descreve com grande precisão o comportamento estatístico de um conjunto muito grande de partículas quânticas (daqui, como disse, o seu enorme sucesso), pretende descrever também o comportamento individual de cada uma das partículas, ou melhor, pretende estabelecer os limites sobre o que podemos entender acerca do comportamento de cada uma das partículas consideradas individualmente. Foi exactamente neste ponto que surgiram enormes divergências entre a comunidade dos físicos. E elas surgiram logo entre os físicos ligados ao aparecimento da mecânica quântica. Se, por um lado estavam os que defendiam a interpretação de Niels Bohr (ele próprio, Werner Heisenberg, Max Born, Wolfgang Pauli, Paul Dirac…), por outro estavam os que se lhe opunham (Albert Einstein, Louis de Broglie, Erwin Schrödinger, Max Planck, Paul Ehrenfest…). Este constituiu o grande "cisma" na física do século XX.

[22] J. Croca, *Towards a Nonlinear Quantum Physics*, World Scientific, 2003.

Aparentemente, o princípio de complementaridade, que constitui a pedra de toque de toda a interpretação bohriana do formalismo quântico ortodoxo, estabelece uma divisão epistemológica entre o mundo macrofísico e o mundo microfísico. No fundo, Niels Bohr estava convencido de que essa divisão não existia. Apenas as características do mundo macrofísico escamoteavam um comportamento que, na sua essência, seria semelhante ao do mundo microfísico. Segundo ele nós estaríamos mal habituados… O mundo macrofísico apenas escamotearia algo que, segundo ele, seriam as nossas dificuldades em entender o mundo à nossa volta. Um mundo, na sua essência, indeterminado. Sabemos, hoje, que Bohr não tinha razão, nem podia ter. Pensar que uma das antinomias kantianas, a antinomia entre livre arbítrio e causalidade podia ser "demonstrada" por uma teoria física, exprimindo-o através de um formalismo matemático, por mais *"lúcido"* que ele aparentasse ser (aos olhos de Bohr evidentemente), apenas revela uma enorme ingenuidade. Não tinha razão pelo que acabámos de expor, mas também porque já foi proposto um novo formalismo, não-linear e causal, mais geral[23], que contém o formalismo quântico ortodoxo como um simples caso particular. Esse formalismo, construído por J. R. Croca e o seu grupo no quadro do programa de investigação científica proposto por Louis de Broglie, por ser muito mais geral, permite-nos entender o sucesso da mecânica quântica bohriana e, consequentemente, os seus limites. Ela não é mais que um caso particular dessa teoria não-linear e causal. Caso particular que corresponde ao estudo do comportamento estatístico de um grande conjunto de partículas, caso em que os métodos estatísticos podem ser aplicados, ou seja, caso em que a aproximação linear pode ser aplicada.

Esse formalismo mais geral rompe com a distinção epistemológica e ontológica entre o mundo macrofísico e o mundo microfísico. Mas não como Niels Bohr pretendeu, impondo o abandono de um qualquer compromisso ontológico, abandono esse que se exprimia no princípio de complementaridade. Um princípio que pretendia mostrar, de uma vez por todas, que tínhamos chegado a um formalismo matemático onde se exprimiria, de forma definitiva, a antinomia kantiana entre continuidade e descontinuidade, ou seja, entre onda e corpúsculo. Conceitos, volto a frisar, a que éramos

[23] *Ibid.*

obrigados a recorrer, mas que não representariam mais que os nossos limites cognitivos.

Falemos, então, desse novo formalismo não-linear e causal que se apoia numa posição realista. Numa posição realista em que os conceitos de onda e de corpúsculo correspondem a algo real que, ao nível da microfísica, estão indissoluvelmente ligados. Daí a necessidade de procurar uma teoria que integre esses dois aspectos num todo coerente. Na descrição dos fenómenos macrofísicos tínhamo-los dissociado. Tínhamo-lo feito porque, aparentemente, esses dois conceitos apareciam disjuntos em macrofísica.

Para introduzir esse novo formalismo, não-linear e causal, é necessário voltar um pouco atrás e falar de um outro processo heurístico.

Falemos de David Bohm. Este físico tem uma importância histórica inegável na física do século XX. Foi ele que, em 1952, mostrou que o teorema de von Neumann[24], um teorema que pretendia ser impossível construir uma interpretação do formalismo quântico ortodoxo alternativa à de Niels Bohr, estava errado. Um dos postulados de que partia não era válido. Foi difícil chegar lá. Demorou 30 anos. Mas bastou dar um contra exemplo. De facto, foi Bohm o primeiro a propor uma interpretação alternativa à de Bohr que tem exactamente a mesma capacidade de descrição dos fenómenos quânticos que esta possuía. Existe, contudo um problema. Bohm preocupou-se tanto em arranjar uma interpretação alternativa do formalismo quântico que descrevesse os mesmos fenómenos que a de Bohr, que manteve intocável esse mesmo formalismo. Manteve a linearidade das equações, ou seja, aceitou o formalismo linear que Bohr interpretara à sua maneira. O resultado foi que nem Bohm, nem todos os que abraçaram a sua interpretação desse formalismo linear, conseguiram conceber uma experiência em que a previsão do resultado fosse diferente da que a interpretação de Bohr previa! Como acabei de dizer, as equações em que Bohm se apoia para descrever os fenómenos quânticos são exactamente as mesmas que Bohr utiliza. São equações lineares. Portanto, optar entre a interpretação de Niels Bohr e a de David Bohm é uma mera questão de gosto filosófico. Mas, apesar de tudo, a interpretação de David Bohm teve o mérito de

[24] J. von Neumann, *Mathematical Foundations of Quantum Mechanics, 1932.*

evidenciar que a interpretação de Niels Bohr não era única possível tal como pretendia o teorema de von Neumann.

Continuemos então com Bohm. Foi ele quem, pela primeira vez, mostrou que podemos decompor a equação de Schrödinger em duas equações. Para isso basta introduzir uma solução muito geral dessa equação de onda, por exemplo, uma onda com a forma $\psi = a\, e^{\frac{i}{\hbar}\varphi}$, e decompor a equação nas suas partes real e imaginária. Este processo levou-o a um sistema de duas equações:

$$\begin{cases} \dfrac{1}{2m}\left(\dfrac{\partial\varphi}{\partial x}\right)^2 + \left(\dfrac{\partial\varphi}{\partial y}\right)^2 + \left(\dfrac{\partial\varphi}{\partial z}\right)^2 + V(x,y,z) - \dfrac{\hbar^2}{2m}\dfrac{\nabla^2 a}{a} = -\dfrac{\partial\varphi(x,y,z)}{\partial t} & (5) \\ \dfrac{1}{m}\nabla\left(a^2\nabla\varphi\right) - \dfrac{\partial a^2}{\partial t} = 0 & (6) \end{cases}$$

A primeira equação é nada mais, nada menos que a equação de Hamilton-Jacobi para a função *S*, agora substituída pela função φ, com a particularidade de possuir um termo adicional

$$Q = -\frac{\hbar^2}{2m}\frac{\nabla^2 a}{a} \qquad (7)$$

David Bohm deu a este termo o nome de *potencial quântico*. Ou seja, temos uma equação semelhante à da mecânica de Newton com um termo adicional. Esta equação descreveria então o carácter local de uma partícula quântica. Mas o *potencial quântico*, segundo Bohm, descreve a interacção entre a partícula quântica e aquilo a que Louis de Broglie, e o próprio David Bohm, chamaram o *meio sub-quântico*. Se existem ondas, e estas são ondas reais, então é necessário encontrar o sujeito do verbo ondular.

A segunda equação também aparece na física construída antes do século XX. É a equação da continuidade. Esta segunda equação que, volto a frisar, está contida na equação de Schrödinger, dá conta do carácter extenso de uma partícula quântica.

Falemos agora de um penúltimo processo heurístico. Ph. Gueret e J. P. Vigier[25] procuraram encontrar uma equação não-linear para substituir a equação de Schrödinger. Pretendiam obter uma equação

[25] Ph. Gueret and J. P. Vigier, in *Lett.Nuovo Cimento*, **38**, 125 (1983).

não-linear que não se afastasse muito da equação de Schrödinger. Para isso escreveram

$$-\frac{\hbar^2}{2m}\left(\frac{\partial^2\psi}{\partial x^2}+\frac{\partial^2\psi}{\partial y^2}+\frac{\partial^2\psi}{\partial z^2}\right)+\gamma\frac{\hbar^2}{2m}\frac{\nabla^2\left(\psi\psi^*\right)^{\frac{1}{2}}}{\left(\psi\psi^*\right)^{\frac{1}{2}}}\psi+V(x,y,z)\,\psi=i\hbar\frac{\partial\psi(x,y,z)}{\partial t} \qquad (8)$$

Esta equação é idêntica à equação de Schrödinger com um termo adicional

$$\gamma\frac{\hbar^2}{2m}\frac{\nabla^2\left(\psi\psi^*\right)^{\frac{1}{2}}}{\left(\psi\psi^*\right)^{\frac{1}{2}}}\psi \qquad (9)$$

que é o *potencial quântico* multiplicado pela função de onda ψ e pelo parâmetro γ. Os seus autores impunham que o parâmetro γ tivesse um valor próximo de zero. Esta imposição pretendia que esta equação não se afastasse muito da equação de Schrödinger. Este truque matemático permitiria, segundo os autores, obter uma equação não-linear que tivesse soluções do tipo solitão (ondas com um perfil estável) dentro do quadro do programa de investigação de Louis de Broglie.

Finalmente, o último processo heurístico de que vos quero falar, foi seguido por J. R. Croca[26], um dos membros do Centro de Filosofia das Ciências da Universidade de Lisboa, e cujos trabalhos se inserem no programa de investigação científica proposto por Louis de Broglie. Baseia-se num claro compromisso ontológico, que é o seguinte: Admitamos que partimos das equações da física construída antes do século XX já nossas conhecidas. Essas equações, a de Hamilton-Jacobi sem o *potencial quântico* e a da continuidade, descrevem características que na física anterior ao século XX apareciam como disjuntas. A primeira descreve o movimento de corpúsculos sujeitos a um potencial clássico *V*. A segunda descreve sistemas extensos. Neste processo, Croca não está a deduzir a equação a que pretende chegar. A equação de que estava à procura é, tal como a equação de Schrödinger, um postulado. Essa equação descreverá sistemas em que as propriedades corpusculares (locais) e ondulatórias (extensas) serão indissociáveis mas, e aqui é que reside a essência desta tentativa,

[26] J. R. Croca, *Towards a Nonlinear Quantum Physics*, World Scientific, 2003.

existirão simultaneamente enquanto características reais das partículas. A equação a que J. R. Croca chegou foi

$$-\frac{\hbar^2}{2m}\left(\frac{\partial^2\psi}{\partial x^2}+\frac{\partial^2\psi}{\partial y^2}+\frac{\partial^2\psi}{\partial z^2}\right)+\frac{\hbar^2}{2m}\frac{\nabla^2\left(\psi\psi^*\right)^{\frac{1}{2}}}{\left(\psi\psi^*\right)^{\frac{1}{2}}}\psi+V(x,y,z)\,\psi=i\hbar\frac{\partial\psi(x,y,z)}{\partial t} \qquad (10)$$

ou seja, exactamente aquela que Gueret e Vigier utilizaram mas em que o parâmetro γ é agora igual à unidade. Outra coisa interessante nesta equação é que só quando o parâmetro γ é igual à unidade é que esta equação se torna verdadeiramente não-linear. Isto foi demonstrado por outro dos colaboradores do nosso centro: A. Rica da Silva[27]. Quando $\gamma \neq 1$ a equação é equivalente a um sistema de equações lineares, isto é, não é uma equação não-linear, contrariamente ao que pretendiam Gueret e Vigier. Importa sublinhar que, após a introdução da equação anterior o formalismo quântico alterou-se, permitindo introduzir toda uma outra interpretação das experiências realizadas e a realizar.

Um ponto importante. Quando aplicamos esta equação à descrição de um grande conjunto de partículas, esta equação é equivalente à equação de Schrödinger, ou seja, podemos aplicar o princípio de sobreposição no tratamento das suas soluções. Isto faz com que o sucesso da mecânica quântica ortodoxa na previsão do comportamento estatístico de um grande conjunto de partículas possa ser integrado, pacificamente, nesta nova descrição. Além disso, todas as soluções da equação de Schrödinger são soluções desta equação, isto é, esta equação dá os resultados que a equação de Schrödinger dá quando aplicada aos exemplos clássicos (oscilador harmónico, átomo de hidrogénio) mas, como não podia deixar de ser, possui outras soluções que não estão contidas na equação de Schrödinger. Isto significa que pode descrever mais minuciosamente o que acontece no mundo microfísico, podendo dar pistas para futuras propostas de experiências como, aliás, já tem acontecido. A equação anterior

[27] A. R. Silva, J. S. Ramos, J. R. Croca and R. N. Moreira, *NonLinear Schrödinger Equation, Burger's Equation and Superposition of Solutions*, in *Causality and Locality in Modern Physics and Astronomy*, eds. S. Jeffers et al. (Kluwer Academic Publishers, 1998).

admite soluções que estão de acordo com o compromisso ontológico de que se parte e que tem a ver com o conceito de partícula quântica que Louis de Broglie defendia desde 1924 até 1927, e que retomou em 1952 após os trabalhos de David Bohm. É importante chamar a atenção para que a equação anterior foi procurada partindo de uma determinada visão do que está por trás das teorias físicas que obedecem ao sexto critério, e apenas a estas, tentando atingir uma descrição unificada do mundo macrofísico e do mundo microfísico, ultrapassando a descontinuidade epistemológica que a mecânica quântica bohriana pretendera impor.

Mais uma vez, na história da ciência, é necessário reunificar dois mundos. Agora já não se trata do mundo sub-lunar e do mundo supra-lunar que Platão classificara como mundo "sensível" e mundo "inteligível", respectivamente. A saturação da informação sensorial sobre o mundo "sensível" não permitiu, no seu tempo, entrever a forma de atingir uma descrição matemática do movimento local nesse mesmo mundo. Um mundo "sensível" onde se pensava não ser possível fazer ciência. Um mundo que resistiria à possibilidade de uma descrição matemática do seu Devir, ou seja, da sua evolução. Por outro lado, o mundo "inteligível" era-o porque conseguiam entrever uma forma de descrever matematicamente o seu Devir. Mas um Devir que se cingia à mera mudança de posição. As limitações dos nossos sentidos tinham simplificado esse mundo, sem que se tivessem apercebido disso. E foi por ter simplificado, inadvertidamente, esse mundo, que Platão ousou propor o primeiro programa de investigação científica que os homens desenvolveram: a descrição matemática (geométrica) dos movimentos (meramente locais) dos seres celestes. A revolução científica do século XVII começou com a unificação desses dois mundos através da proposta de Copérnico. Galileu apercebeu-se das consequências da adopção dessa proposta e defendeu-a veementemente, pois ela permitia-lhe usar, para os movimentos locais à superfície da Terra, o mesmo método, conscientemente desta vez, que os antigos tinham utilizado para descrever os movimentos locais celestes. A unificação ontológica que o modelo de Copérnico impunha, arrastava consigo uma unificação epistemológica.

Tal como na época de Galileu, é necessário lutar, mais uma vez, por uma nova unificação de dois mundos. O mundo macrofísico que a física estudou desde a revolução científica até ao início do século XX,

e o mundo microfísico com que se defrontou, conscientemente, a partir daí. Mas existe uma diferença subtil entre o que se passou no período greco-helenístico e no início do século XX. Na antiguidade, as limitações dos nossos sentidos tinham simplificado o mundo supra-lunar tornando-o "inteligível", ou seja, passível de uma descrição matemática. No início do século XX, foram novamente as limitações dos nossos sentidos, que se manifestam quando observamos fenómenos que estão associados a um mundo também ele muito distante, à sua maneira, que levaram alguns físicos a convencer-se que deveríamos cingir-nos a uma descrição meramente matemática dos fenómenos observados, sem procurar nenhuma correspondência com uma realidade subjacente. Na antiguidade foram as limitações dos nossos sentidos que nos fizeram considerar esse mundo como "inteligível", ou seja, passível de uma descrição matemática. Pelo contrário, foi a dificuldade de tornar inteligível o que se observava no mundo microfísico que levou à adopção de uma posição instrumentalista. A mecânica quântica ortodoxa seria, assim, um instrumento em que estaria representada, não uma realidade subjacente aos fenómenos que pretendíamos descrever com ela, mas, acima de tudo, a nossa dificuldade de entender o mundo microfísico. Um instrumento em que estava, agora expressa essa dificuldade de uma *forma matematicamente lúcida*. Uma *forma matematicamente lúcida* que acolhia uma interpretação apoiada no princípio de complementaridade de Höffding-Bohr.

O último critério de validação de uma teoria permite-nos defender, de uma forma mais coerente do que o tentaram fazer Karl Popper, Imre Lakatos e Larry Laudan, uma visão continuista do avanço da ciência, contra uma visão descontinuista desse mesmo avanço, tal como o defendeu Thomas Kuhn ao defender a incomensurabilidade entre teorias científicas e, por conseguinte, porque é o caso que me interessa, entre teorias físicas. Paul Feyerabend radicalizou essa proposta kuhniana adoptando uma posição radicalmente céptica. A única atenuante de Feyerabend foi ter-se confrontado com uma física, a física dominante no século XX, que baixou a guarda contra as tendências irracionalistas, ao aceitar o princípio de complementaridade como pedra de toque de toda a interpretação do formalismo quântico ortodoxo.

Para encontrar uma forma de combater estas posições, devemos "filtrar o ruído" criado por diversos projectos de teoria que a física

tem vindo a construir ao longo do seu já longo trajecto. Se nos cingirmos às teorias que conseguiram ultrapassar o último critério, podemos vislumbrar um fio condutor que nos dá maior esperança de ir levantando mais *uma ponta do grande véu* que cobre uma realidade subjacente aos fenómenos, tal como o afirmava Albert Einstein na resposta a Paul Langevin, quando este lhe solicitou uma apreciação à tese que Louis de Broglie tinha acaba de apresentar na Universidade de Paris, em que propunha que a todas as partículas materiais estaria associada uma onda. Seguindo este fio condutor, somos capazes de eliminar a *incomensurabilidade* entre teorias. Cingindo-nos a este tipo de teorias verificamos que é possível estabelecer a sua *comensurabilidade*. Mas uma comensurabilidade no sentido atrás referido, ou seja, que a antiga *boa* teoria está, toda ela, contida na nova boa teoria. É seguindo este fio condutor que somos obrigados a confrontar-nos com o grande problema não resolvido pela física do século XX: o problema do *dualismo onda-corpúsculo*. É na resolução deste problema no quadro do espaço e do tempo e não abdicando da categoria de causalidade que poderemos recuperar um entendimento do mundo que lança pontes entre a *boa*[28] física construída antes do século XX, a *boa* física construída durante o século XX, e lançar pontes para a *boa* física a construir no futuro. Mas para se poder dizer isto é indispensável que conheçamos profundamente, não apenas a história da física, não apenas os grandes problemas filosóficos que lhe estão associados, mas combater na própria fronteira em que a física se encontra hoje a combater, isto é, conhecer profundamente o que os físicos fizeram ao longo da história, conhecer profundamente os problemas epistemológicos com que ela se defrontou no passado, e conhecer profundamente a física dos nossos dias. Os historiadores da ciência têm um papel importantíssimo a desempenhar. Sem a história da ciência não é possível pensar sobre ela e fazer filosofia da ciência. Também nunca foi possível fazer filosofia da física, de forma profícua, sem conhecer profundamente a própria física. O filósofo da ciência tem de "sujar as mãos" na própria ciência, isto é, praticá-la, porque só assim evita a situação, infelizmente demasiado comum, de andar reverentemente a ouvir o que os cientistas têm para lhe dizer. Na maior parte dos casos aceitando-o como certo. Um filósofo da ciência

[28] Constituída pelas teorias que obedecem aos 6 critérios enunciados no início deste texto.

tem de ser capaz de, num determinado domínio científico, falar de igual para igual com os homens que fazem ciência nesse domínio.

Regressemos ao programa de investigação científica proposto por Louis de Broglie. Na figura 1 pretende-se evidenciar, de uma forma inevitavelmente tosca, o compromisso ontológico assumido por Louis de Broglie sobre o que poderá ser uma partícula quântica. Mas esta figura é uma representação de uma função do tipo

$$\varphi=\sqrt{\frac{\hbar^2}{\pi\sigma_0}\left[e^{-\frac{(kx-\omega t-\varepsilon_0)^2}{2\sigma_0{}^2}}+\alpha e^{-\frac{(kx-\omega t-\varepsilon)^2}{2\sigma^2}}\right]}e^{i(kx-\omega t)} \qquad (11)$$

que é uma solução particular da equação não-linear (10) proposta por J. R. Croca.[29] Nesta solução particular existe uma aproximação provisória que é a adição da representação das características corpusculares (primeira parcela da soma) com a representação das características ondulatórias (segunda parcela da soma). A interacção entre as duas será necessariamente mais complexa que a representada pala sua soma.

A figura 1 tenta representar uma fortíssima concentração de energia representada pela zona em que a onda possui uma maior amplitude, que será responsável pelos fenómenos em que as propriedades corpusculares se evidenciam, e uma onda, muito mais extensa, que será responsável pelos fenómenos em que as propriedades ondulatórias se evidenciam. Mas, volto a frisar, esta é apenas uma paupérrima imagem do que, à escala em que estamos a descrever os fenómenos, admitimos poder ser uma partícula quântica. Este é o compromisso ontológico, claro que provisório, porque um compromisso ontológico é sempre provisório, que nos faz recuar aos primórdios da física quântica, quando Max Planck (1900) e Louis de Broglie (1924) introduziram as duas equações fundamentais,

$$E=h\nu \qquad \text{e} \qquad \vec{p}=h\vec{k}\,,$$

A figura 1 tenta representar uma fortíssima concentração de energia representada pela

[29] J. R. Croca, *Towards a Nonlinear Quantum Physics*, World Scientific, 2003, p. 87.

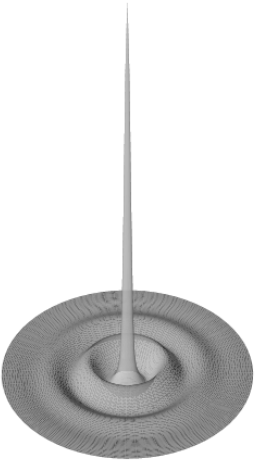


Figura 1

zona em que a onda possui uma maior amplitude, que será responsável pelos fenómenos em que as propriedades corpusculares se evidenciam, e uma onda, muito mais extensa, que será responsável pelos fenómenos em que as propriedades ondulatórias se evidenciam. Mas, volto a frisar, esta é apenas uma paupérrima imagem do que, à escala em que estamos a descrever os fenómenos, admitimos poder ser uma partícula quântica. Este é o compromisso ontológico, claro que provisório, porque um compromisso ontológico é sempre provisório, que nos faz recuar aos primórdios da física quântica, quando Max Planck (1900) e Louis de Broglie (1924) introduziram as duas equações fundamentais,

$$E = h\nu \qquad \text{e} \qquad \vec{p} = h\vec{k}\,,$$

nas quais estavam indissoluvelmente ligadas as características corpusculares e ondulatórias. Características reais (actuais) e não meramente potenciais como a interpretação de Bohr do formalismo quântico pretendeu impor. Este recuo não é inócuo, pois foi neste quadro que a equação não-linear atrás referida foi atingida. Além disso, foi dentro deste quadro que têm sido propostas experiências que pretendem testar a plausibilidade deste compromisso ontológico. Experiências em que a previsão do resultado é significativamente diferente da que a interpretação bohriana do formalismo quântico ortodoxo prevê. Desta forma, afirmar que é equivalente assumir uma posição realista, ou uma posição instrumentalista, revela-se

profundamente errada. Admitir a existência de uma estrutura real de uma partícula quântica levou alguns físicos a propor a realização de experiências que tentam evidenciar, mesmo que de forma indirecta, a existência dessa estrutura[30].

A aceitação da equação não-linear (10) impõe um corte ontológico e epistemológico que importa explicitar. A *forma matematicamente lúcida* que Bohr referia na frase que citei atrás é, nada mais, nada menos, que a análise de Fourier. O *princípio de complementaridade* de Bohr atribui a esta análise um estatuto ontológico e epistemológico privilegiado. Como sabemos, esta análise diz-nos que é possível analisar e sintetizar qualquer função $\boldsymbol{\Psi}$, por exemplo, uma solução de uma qualquer equação linear da mecânica quântica ortodoxa, utilizando as funções seno e coseno que constituem a base de um espaço funcional a que damos o nome de espaço de Hilbert. Além disso, só estas funções, as funções que constituem essa base, podem possuir uma frequência bem definida. No caso geral, a função $\boldsymbol{\Psi}$ é representada pela soma de um conjunto infinito, discreto ou contínuo, de senos e cosenos, cada um com a sua frequência. Desta situação resulta que quando essa função $\boldsymbol{\Psi}$ pretende representar uma única partícula, e um dos postulados da mecânica quântica bohriana diz que nessa função está contida toda a informação que podemos possuir sobre essa partícula, é-nos impossível dizer qual a energia que a partícula possui (em acto) antes de uma qualquer operação de medida. De facto, como $E = h\nu$ e**,** portanto, a cada frequência ν está inevitavelmente associada uma energia *E* diferente, deveríamos concluir que a partícula possuiria infinitos valores para a energia.

---

[30] - J. and M. Andrade e Silva, C. R. Acad. Sci. (Paris) 290 (1980) 501.
- F. Selleri, Ann.Found. L. de Broglie, 4 (1982) 45.
- F. Selleri, Found. Phys. 12 (1983) 1087; 17 (1987) 739.
- A. Garuccio, V. Rapizarda and J. P. Vigier, Phys. Lett. A 90 (1982) 17.
- J. R. Croca, in *Microphysical Reality and Quantum Formalism*, eds. A. Van der Merwe, F. Selleri, G, Tarozzi, Kluwer, Dordrecht, 1988.
- J. R. Croca, Ann. Found. L. de Broglie, 14 (1989) 323.
- J. R. Croca, in Waves and Particles in Light and Matter, eds. Alwyn Van der Merwe and A. Garuccio, Plenum, 1994.
- J. R. Croca, A. Garuccio, V. Lepori e R. N. Moreira, Found. of Phy. Lett. 3 (1990) 557.
- J. R. Croca, R. Moreira and A. J. Rica da Silva, in *Causality and Locality in Modern Physics*, eds. G. Hunter et al., Kluwer Academic Publishers, 1998.

Como em todas as operações de medida detectamos sempre um, e só um valor para a energia da partícula, a conclusão anterior não é válida. Bohr "resolveu" o problema afirmando que esses valores da energia da partícula contidos na função de onda **Ψ** não existem em acto mas apenas em potência. Ou seja, que o formalismo quântico, interpretado por Bohr, atribui um papel determinante ao acto de medir, levando alguns a concluir que é o observador que, no acto de medir, cria o próprio objecto que mede. Foi esta posição bohriana que levou ao grande "cisma" na física do século XX. Quando aplicamos o formalismo quântico ortodoxo ao conjunto de um grande número de partículas, podemos sempre interpretá-lo como uma descrição estatística da distribuição de energias pelas partículas que constituem esse conjunto. Mas cada uma delas possuiria uma energia bem definida. Tal como na distribuição das alturas de um conjunto de indivíduos, existirão várias alturas diferentes, mas tendo cada um dos indivíduos uma única altura. Desta forma, quando aplicamos o formalismo quântico ortodoxo a um grande conjunto de partículas, não existe nenhum problema. Podemos utilizar equações lineares, como primeira aproximação, e nenhum problema epistemológico se levanta. Mas, como a mecânica quântica ortodoxa pretende ser uma teoria completa e, dado que as equações da mecânica quântica ortodoxa são lineares, o que afirmamos para um grande conjunto de partículas deve aplicar-se a uma só partícula, ou a um número reduzido de partículas em que a abordagem estatística já não é válida. Bohr pensou, ingenuamente, que esse formalismo seria a tal *forma matematicamente "lúcida"* em que as dificuldades que sempre sentíramos na descrição dos fenómenos estariam finalmente representadas. Esse formalismo representaria para Bohr, como disse antes, acima de tudo as nossas limitações cognitivas. Esse formalismo não pretenderia mais do que descrever a forma como apreendemos os fenómenos, assumindo, assim, uma concepção instrumentalista da teoria quântica. Existem ainda hoje muitos que interpretam erradamente o significado profundo da mecânica quântica ortodoxa. Todos os desenvolvimentos posteriores, que tenham aceitado os postulados da mecânica quântica ortodoxa, não deveriam ter esquecido o significado profundo que está na base da interpretação bohriana do formalismo quântico ortodoxo. O espaço de Hilbert e as funções que constituem a base desse espaço, ou seja, os senos e os cosenos, no qual construímos a nossa descrição do mundo microfísico,

teriam muito mais a ver, segundo Bohr, com as nossas limitações cognitivas, do que com um mundo que esteja para lá da forma como o apreendemos. Os que aceitam essa interpretação deveriam há muito ter deixado de falar de um mundo para além da nossa apreensão dele. É ridículo ouvirmos hoje dizer que as experiências que se estão a realizar com o LHC vão "mostrar-nos" como o universo foi criado. Como se a interpretação de Bohr do formalismo quântico ortodoxo, que nenhum dos que fazem essa afirmação pôs em causa, lhes permitisse falar de um universo para além do que nós, com as nossas limitações cognitivas, apreendemos dele. Como se vê a física chegou à confusão total.

É necessário chamar ainda a atenção para outra consequência da atribuição de um estatuto epistemológico privilegiado à análise de Fourier, como o faz a interpretação bohriana do formalismo quântico ortodoxo. Os senos e cosenos, que constituem a base do espaço de Hilbert onde se descreve a forma como apreendemos o mundo à nossa volta, são funções que têm uma amplitude constante em todo o domínio do seu argumento, ou seja, em todo o espaço e em todo o tempo[31]. São funções que existiriam desde a origem dos tempos até ao seu fim, se é que faz sentido falar destes limites, e que existiriam por todo o espaço! Não é de estranhar que alguns físicos tenham tentado, e tentem ainda hoje, idealizar e realizar experiências em que pretendem evidenciar a existência de retroacções no tempo![32] Claro que essas mesmas experiências já foram explicadas, dentro do programa de investigação proposto por Louis de Broglie, sem recorrer a tão bizarras conclusões[33], mostrando que, afinal, podemos explicá-las sem

[31] J. R. Croca, *Towards a Nonlinear Quantum Physics*, World Scientific, 2003, pp. 8-13, 62-65 e 90-93.

[32] - J. A. Wheeler, *Delayed choice experiments,* in Proceedings of New Orleans Conference on the Mathematical Foundations of Quantum Theory, ed. A. Marlow, Academic Press (New York, 1978);
- Kaiser, R. Clothier, S. A. Werner, H. Rauch and W+olwitch, Phys. Rev. A 45 (1992) 31;
- H. Rauch, Il Nuovo Cimento, B110, (1995) 557;
- V. Jacques, E. Wu, F. Grosshans, F. Treussart, P. Grangier, A. Aspect and J.-F. Roch, arXiv :quant-ph/0610241v1 28Oct2006.

[33] - J. R. Croca, R. Moreira and A. J. Rica da Silva, in *Causality and Locality in Modern Physics*, eds. G. Hunter et al., Kluwer Academic Publishers, 1998;

abandonar as concepções de espaço, tempo e causalidade, ou seja, mantendo-nos no quadro da racionalidade a que a ciência nos tinha habituado. Tudo isto não faz mais que tornar evidente a enorme confusão em que a Física tem vivido desde que a interpretação bohriana do formalismo quântico ortodoxo foi aceite pela generalidade da comunidade dos físicos.

Voltemos à física realista. Numa equação não-linear, em geral, o todo não é igual à soma das partes. Desta forma, o que afirmamos para um grande conjunto de partículas, já não se pode aplicar a cada uma das partículas isoladamente. Para isso, é necessário retirar o estatuto ontológico e epistemológico privilegiado que a análise de Fourier possui na interpretação bohriana do formalismo quântico ortodoxo. O corte epistemológico reside em afirmar, contrariamente ao que nos diz a análise de Fourier, que uma onda que não seja representada por um seno ou um coseno pode ter uma frequência bem definida. É isto que J. R. Croca defende.[34] Uma das consequências da não-linearidade e da recusa de atribuir à análise de Fourier um estatuto ontológico e epistemológico privilegiado é o abandono do *princípio de complementaridade* e das relações de indeterminação de Bohr-Heisenberg (impropriamente conhecido como princípio de incerteza de Heisenberg) que são uma consequência daquele. No quadro da, muito mais geral, análise de onduletas (wavelet analysis), desenvolvida mais acentuadamente desde os anos oitenta do século XX, a análise de Fourier pode ser considerada como um mero caso particular. Nesse quadro é possível deduzir relações mais gerais que as relações de indeterminação de Bohr-Heisenberg.[35] Eis as relações a que J. R. Croca chegou:

- J. E. Araújo, J. L. Cordovil, J. R. Croca and R. N. Moreira, *A Causal and Local Way to Interpret the Experimental Realization of Wheeler's Delayed-Choice Gedanken Experiment*, in print;
- J. E. Araújo, J. L. Cordovil, J. R. Croca, R. N. Moreira and A. Rica da Silva, *Some Experiments that Question the Completeness of Orthodox Quantum Mechanics*, in print;

[34] J. R. Croca, *Towards a Nonlinear Quantum Physics*, World Scientific, 2003, p. 92.

[35] *Ibid.,* pp. 104-105.

$$\Delta x^2 = \frac{h^2}{\Delta p_x{}^2 + \frac{h^2}{\sigma_0{}^2}} \qquad (12)$$

Estas relações foram atingidas por J. R. Croca[36] a partir de uma visão realista de partícula quântica e utilizando uma análise mais geral que a análise de Fourier, a análise de onduletas de tipo Morlet, em que σ_0 representa as dimensões da onduleta base. As relações que Croca propõe transformam-se mas relações de Bohr-Heisenberg

$$\Delta x \, \Delta p_x = h$$

quando as dimensões da onduleta base de Morlet tendem para infinito, ou seja, quando essas onduletas se identificam com os senos ou os cosenos da análise de Fourier.

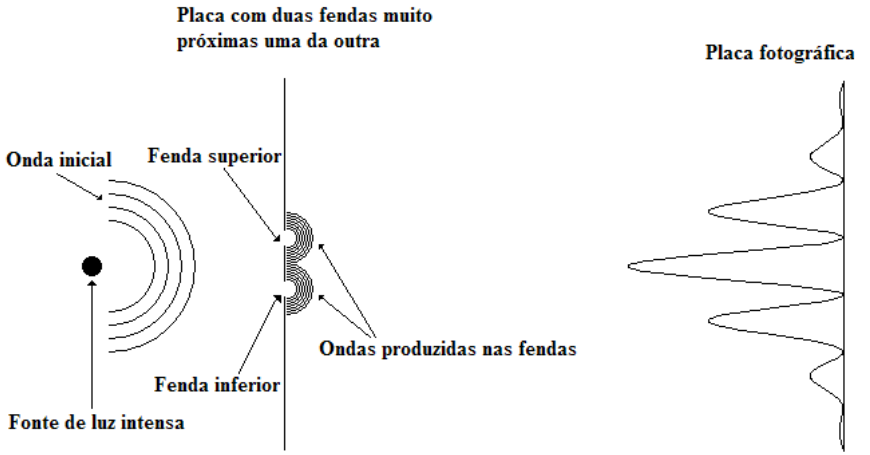
Isto é tanto mais interessante quando já foi evidenciado que o limite imposto pelas relações de dispersão de Bohr-Heisenberg é ultrapassado em experiências já realizadas. Com microscópios de varrimento de campo próximo, a previsão da posição e da velocidade de uma partícula, num futuro acto de medida, é conhecida com uma precisão mais de 10 vezes superior à que essas relações pretendiam impor.[37]

Falemos agora de um tipo de experiência que serve para confrontar os dois programas de investigação científica de que temos vindo a falar. Numa experiência de interferências de tipo Taylor criamos as condições para que no dispositivo interferencial exista apenas uma única partícula de cada vez. Este tipo de experiências tem sido realizado com diversos tipos de partículas. Tanto podemos falar de neutrões como de fotões, apesar de os neutrões serem partículas de tipo fermiónico e os fotões serem partículas de tipo bosónico. As primeiras são mais fáceis de isolar de acordo com o princípio de exclusão de Pauli. Mas as segundas também têm sido utilizadas desde que L. Mandel criou fontes com cristais não lineares, que utilizam a chamada *parametric down-conversion*. Estas constituem fontes monofotónicas fidedignas. Em experiências deste tipo emergem os problemas epistemológicos fundamentais associados à controvérsia

[36] *Ibid.*, p. 104.
[37] *Ibid.*, p. 125.

sobre os fundamentos da mecânica quântica ortodoxa. Temos aqui um caso em que queremos aplicar o formalismo quântico ortodoxo ao estudo do comportamento de uma só partícula.

Comecemos, no entanto, por referir o que se observa numa experiência de Young, no quadro da teoria electromagnética de Maxwell, em que utilizamos uma fonte de luz intensa. Neste caso enviamos para o interferómetro uma enorme quantidade de fotões simultaneamente. Vamos representá-la na figura 2. Temos uma fonte intensa de luz, à frente da qual existe um divisor de frente de onda, ou seja, duas fendas numa placa que dividem a frente de onda circular criada pela fonte em duas frentes de onda cilíndricas, que saem de cada uma das duas fendas. Estas duas frentes de onda cilíndricas vão sobrepor-se, ou seja, vão interferir entre si dando origem ao aparecimento de uma sucessão de zonas claras e escuras numa placa fotográfica apropriadamente colocada a uma certa distância dessas fendas. Esta sucessão de zonas claras e escuras é facilmente explicada admitindo que a luz é um fenómeno ondulatório. As zonas claras



Interferómetro de Young

Figura 2

correspondem às regiões da placa fotográfica em que a "crista" da onda que saiu da fenda superior se sobrepõe à "crista" da onda que saiu da fenda inferior reforçando a intensidade da luz. As zonas escuras correspondem às regiões da placa fotográfica em que a "cava"

da onda que saiu da fenda superior se sobrepõe à "crista" da onda que saiu da fenda inferior diminuindo a intensidade da luz e vice-versa.
Se em vez da fonte de luz tivéssemos utilizado um emissor de corpúsculos newtonianos, por exemplo, uma metralhadora na posição de tiro contínuo nas mãos de um atirador. Neste caso, as balas sairiam com uma direcção aleatória em torno de uma posição alinhada com um ponto entre as duas fendas. As fendas teriam de ser redimensionadas para deixarem passar estes pequenos projécteis sempre que eles não colidissem com a

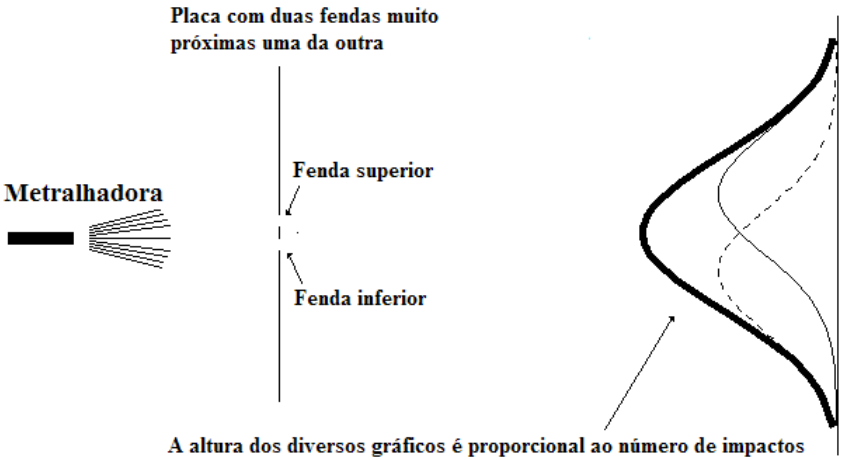


Figura 3

placa que contém as duas fendas. Neste caso, observaríamos algo completamente diferente. A figura 3 tenta representar três situações: a distribuição dos impactos se só a fenda superior estivesse aberta (curva cheia fina), se só a fenda inferior estivesse aberta (curva a tracejado) e a soma destas que representaria a distribuição dos impactos quando as duas fendas estivessem abertas (curva cheia espessa). Devemos concluir que o resultado esperado é diferente consoante a característica do objecto que passa pelas duas fendas é eminentemente ondulatória (primeiro caso) ou eminentemente corpuscular (segundo caso). No primeiro caso a onda passou pelas duas fendas simultaneamente. No segundo caso, cada projéctil passou apenas por uma das duas fendas. No primeiro caso obteríamos uma figura interferencial. No segundo caso obteríamos uma distribuição gaussiana ou normal. Isto é o que se observaria em situações em que a

física construída antes do século XX se podia aplicar, ou seja, estaria de acordo com o electromagnetismo maxweliano no primeiro caso, ou de acordo com a mecânica newtoniana no segundo caso.

Passemos, finalmente, para uma experiência de interferências de tipo Taylor, em que enviamos apenas um fotão de cada vez. Admitamos que a taxa média de emissão de fotões é de um por segundo. Poderíamos fazer uma revelação da placa fotográfica ao fim de 10 segundos e observaríamos uma distribuição aparentemente caótica de 10 pontos luminosos nessa placa. Ao fim de 20 segundos e observaríamos uma distribuição ainda aparentemente caótica de 20 pontos luminosos nessa placa. Mas à medida que fossemos aumentando o tempo de exposição, a distribuição de pontos luminosos ia-se cada vez mais aproximando da distribuição obtida quando, no caso considerado há pouco, tivéssemos enviado muitos fotões simultaneamente, ou seja, da distribuição que o electromagnetismo de Maxwell tão bem explicava recorrendo apenas ao conceito de onda. Poderíamos assim concluir que cada um dos fotões passou simultaneamente pelas duas fendas como se fosse uma onda e, por isso, interferiu consigo próprio acabando por reproduzir, ao fim de muitos eventos semelhantes, a distribuição interferencial habitual.

A próxima situação é preparada porque queremos receber informação sobre a passagem do fotão, enquanto partícula, pelas duas fendas. Colocamos um detector após cada uma das fendas ligados a um contador de coincidências que, como o seu próprio nome indica, apenas contaria os eventos em que o fotão passasse pelos dois orifícios simultaneamente. O que se verifica é que o número de eventos deste tipo é nulo, ou seja, o fotão nunca passou pelos dois orifícios simultaneamente. Chegamos assim a um aparente paradoxo.

- No primeiro caso, o fotão passou por uma das fendas

e

pela outra simultaneamente, para que seja possível explicar a formação de uma figura de interferências recorrendo ao conceito de onda.

- No segundo caso, o fotão passou por uma das fendas

ou

pela outra. Nunca pelas duas simultaneamente. Neste caso, o fotão evidencia as suas características corpusculares.

Niels Bohr concluiu que estávamos a ser confrontados com uma situação em que, finalmente e definitivamente, as nossas limitações cognitivas emergiam. As categorias de continuidade e descontinuidade, ou seja, de onda e corpúsculo, evidenciariam as suas limitações. São categorias necessárias mas inconciliáveis, e o formalismo que descreve os fenómenos quânticos exprime este *resíduo irracional* de uma *forma matematicamente lúcida.* Para Bohr, a mecânica quântica ortodoxa viria, desta forma, fornecer-nos um caso concreto que traduziria a antinomia kantiana entre uma visão continuista e uma visão descontinuista do mundo que, por sua vez, estaria relacionada com a nossa incapacidade de, simultaneamente, descrever o que observamos no quadro do espaço e tempo por um lado, e relacionar causalmente os fenómenos observados, por outro. E aqui está a relação de complementaridade que Höffding teria detectado entre as funções psicológicas ver e compreender. Para isso, Bohr defendeu que o máximo que poderíamos dizer sobre este fenómeno, era que a partícula tinha passado *potencialmente* pelas duas fendas, com uma determinada probabilidade *P* de passar pela fenda inferior, e uma probabilidade *(1-P)* de passar pela fenda superior. Para isso a onda que descreveria o comportamento da partícula perderia a sua relação realista com uma onda física, e passaria a ser interpretada apenas como uma onda de probabilidade. Uma onda meramente matemática. Cá estamos, mais uma vez, a constatar como a atitude instrumentalista foi introduzida na física do século XX, neste caso através da mecânica quântica ortodoxa e através do princípio de complementaridade que está na base da interpretação bohriana do formalismo quântico ortodoxo. Quando deixamos as duas fendas abertas estamos a evidenciar o carácter ondulatório de uma partícula quântica, porque a onda de probabilidade, associada à partícula, passou potencialmente, e simultaneamente, por ambas as fendas dando origem a uma figura interferencial. Quando colocamos detectores imediatamente após as duas fendas para saber por onde a partícula passou, verificamos que ela apenas passou por uma delas, deixando, nesta configuração experimental, de poder provocar, como é evidente, uma figura interferencial. Estamos neste caso a esbater, se não mesmo a eliminar, as características ondulatórias da partícula quântica, e a evidenciar as suas características corpusculares. Importa agora chamar a atenção para um postulado da interpretação ortodoxa da mecânica quântica. Sendo as duas ondas que passam por cada uma das fendas

meras ondas de probabilidade, quando se detecta a passagem do fotão por uma das fendas, imediatamente a probabilidade de o fotão ter passado por essa fenda se torna igual a um, ou seja, passa a ser o acontecimento certo, enquanto a probabilidade de o fotão ter passado pela outra fenda passa imediatamente a ser igual a zero, ou seja, passa a ser um acontecimento impossível. Sendo estas ondas meras ondas matemáticas, ou seja, ondas de probabilidade, de acordo com a interpretação ortodoxa nada terá passado pela fenda em que essa onda de probabilidade se anulou. Exactamente porque essa onda é um mero artifício matemático que não corresponde a nenhum objecto físico real. Esta interpretação é-nos imposta pelo quinto postulado da mecânica quântica ortodoxa, o postulado da redução instantânea do trem de ondas. No acto de medida de que estamos a falar, das duas ondas de probabilidade que nos davam a probabilidade de a partícula passar por uma das fendas ou pela outra, após a detecção da partícula numa das fendas, a onda de probabilidade, que nos davam a probabilidade de a partícula passar pela outra fenda anulou-se instantaneamente. É este postulado que, volto a sublinhar, leva alguns a dizer que, na interpretação de Bohr do formalismo quântico ortodoxo, é o observador que, no acto de medição, transforma em acto uma das duas potencialidades, ou seja, que é o próprio observador que cria a observação. Volto a realçar que, para Bohr e para Höffding toda esta estranha descrição tem mais a ver com as nossas dificuldades cognitivas do que com qualquer outra coisa. O formalismo quântico ortodoxo, e a interpretação bohriana deste, representariam muito mais a nossa forma de observar o mundo do que esse mundo.

No quadro do programa de investigação realista de Louis de Broglie a explicação não evidencia nenhum *resíduo irracional*. Relembremos a figura 1 onde está representada a imagem, mais uma vez realço, tosca de uma partícula quântica. Nesta concepção de uma partícula quântica esta assume, simultaneamente, uma forte concentração de energia que lhe dá um carácter muito localizado que representa aquilo a que chamamos corpúsculo ("descontinuidade") e uma onda com uma energia muitíssimo inferior que lhe dá o carácter extenso ("continuidade"). A partícula é este todo em que o corpúsculo jamais pode existir isolado de uma onda física. Já o inverso não será verdadeiro. Pode existir uma onda sem corpúsculo. A partícula quântica é, neste compromisso ontológico, um ser com estrutura. Uma estrutura que se mantém através de uma relação complexa entre

corpúsculo e onda. Uma estrutura em que a relação entre a mecânica newtoniana e o electromagnetismo maxweliano pode ser reavaliada. Toda a matéria, ou seja, todo o objecto da Física deve passar a ser considerado como possuindo uma estrutura simultaneamente corpuscular e ondulatória tal como o propuseram de início Max Planck e Louis de Broglie. É esta posição que leva todos os que se integram no programa de investigação científica proposto por Louis de Broglie a procurar estabelecer pontes entre o mundo macrofísico e o mundo microfísico. Sem a ingenuidade daqueles que, na Física construída antes do século XX, acreditavam que o seu compromisso ontológico era uma imagem fidedigna da coisa em si, nem a ingenuidade daqueles que, na física construída durante o século XX, caíram no posição oposta ao defenderem que devíamos, definitivamente, abandonar qualquer tentativa de criar compromissos ontológicos, mesmo que provisórios, face a uma realidade subjacente. Uma realidade subjacente que deveríamos, definitivamente, afastar das nossas preocupações ao fazer física.

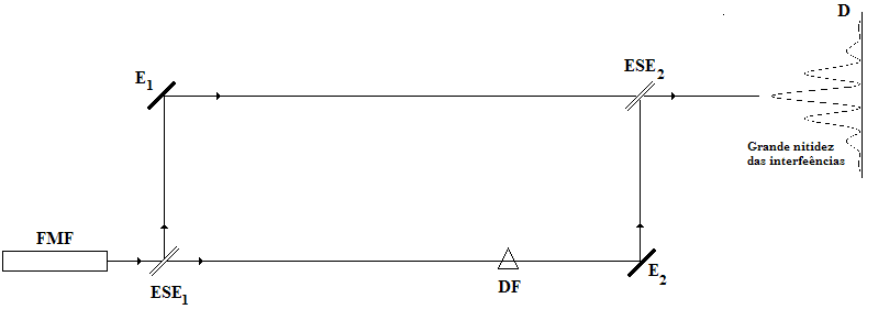
Regressemos à experiência de interferências. No quadro do programa de investigação científica proposto por Louis de Broglie não existe nenhum paradoxo. O corpúsculo, ou seja, a concentração de energia fortemente localizada apenas pode passar por uma das fendas. A onda extensa, física e não apenas matemática, que o acompanha, terá passado pelas duas. Como apenas a forte concentração de energia pode ser detectada directamente pelos detectores usuais, explica-se a razão porque o corpúsculo (na descrição ortodoxa falar-se-ia de partícula) nunca passa pelas duas, quando utilizamos um dispositivo experimental que pretende evidenciar o carácter corpuscular de uma partícula quântica. Mas agora uma pergunta se coloca inevitavelmente. Como podemos nós evidenciar que a onda passou pelas duas fendas? Directamente não o conseguimos fazer porque a energia associada à onda que não transporta o corpúsculo e que passa pela outra fenda não possui energia suficiente para que um detector usual a detecte. Mas, de acordo com o programa realista de Louis de Broglie, não se consegue explicar a formação da figura de interferências, numa experiência com um único fotão no interferómetro de cada vez, sem admitir que é a sobreposição da onda que transporta o corpúsculo com a que não o transporta e que passou pela outra fenda que, de alguma forma, o vai guiar criando zonas mais prováveis e zonas menos prováveis em que esse corpúsculo vai

impressionar a placa fotográfica, dando origem, após um grande número de impactos, a uma figura interferencial. Neste compromisso ontológico convencionou-se chamar onda vazia à onda desprovida de corpúsculo e representá-la pela letra $\boldsymbol{\theta}$. Claro que a onda que acompanha o corpúsculo é também uma onda física do mesmo tipo. Por essa razão também a representamos pela letra $\boldsymbol{\theta}$, apesar de não ser uma onda vazia. De acordo com esta visão realista de uma partícula quântica é a interferência destas duas ondas físicas, a onda $\boldsymbol{\theta}$ que acompanha o corpúsculo e a onda $\boldsymbol{\theta}$ vazia que não o transporta, que origina o fenómeno de interferências.

Podemos agora recolocar a questão levantada imediatamente acima. Será possível conceber experiências em que se evidencie a realidade física destas ondas $\boldsymbol{\theta}$? Para que a comunidade dos físicos possa aceitar a existência real destas ondas, abandonando a interpretação ortodoxa, é necessário que se concebam experiências em que o resultado previsto, no quadro do programa de investigação científica proposto por Louis de Broglie, seja diferente do que prevê o programa de investigação proposto por Niels Bohr. Se tal não fosse possível estaríamos numa situação semelhante à que acontece entre a interpretação de Bohr e a interpretação de Bohm, em que não conseguimos conceber experiências em que a previsão do resultado seja diferente entre as duas. A resposta à pergunta colocada no início deste parágrafo é afirmativa. De facto, existem hoje muitas propostas de experiências em que o resultado previsto é diferente, consoante adoptemos a interpretação ortodoxa ou a interpretação de Louis de Broglie desenvolvida por J. R. Croca e o seu grupo de investigação. Vou referir apenas uma.

Imaginemos um interferómetro de Mach-Zehnder onde se produza um padrão de interferências. Esse tipo de interferómetro está representado na figura 4. A fonte monofotónica (***FMF***) envia um fotão de cada vez para o primeiro espelho semi-espelhado ($\boldsymbol{ESE_1}$). Num espelho deste tipo, parte da onda é reflectida para o espelho $\boldsymbol{E_1}$ e depois deste é reflectida para o segundo espelho semi-espelhado ($\boldsymbol{ESE_2}$), enquanto a outra parte da onda é transmitida no $\boldsymbol{ESE_1}$ para o espelho $\boldsymbol{E_2}$ onde é reflectida para o segundo espelho semi-espelhado ($\boldsymbol{ESE_2}$). É sobre este segundo espelho semi-espelhado que as ondas que seguem as duas vias diferentes, se vão sobrepor e, consequentemente, originar o padrão de interferências observado no detector ***D***. Falta só explicar que o papel representado pelo desviador

de fase (***DF***) é apenas o de fazer com que nenhuma onda siga o trajecto complementar no segundo espelho semi-espelhado (***ESE₂***), ou seja, de acordo com a orientação da figura, não siga na vertical para cima.



Interferómetro de Mach-Zehnder

Figura 4

Este tipo de interferómetro reproduz uma figura interferencial, tal como o interferómetro de tipo Young de que falámos atrás, após um grande número de fotões ter atingido o detector ***D***, que pode ser uma placa fotográfica. Isto porque, também neste caso, a partir de uma frente de onda inicial se produzem duas frentes de onda no primeiro espelho semi-espelhado (***ESE₁***). Estas duas frentes de onda, tal como no interferómetro de Young, voltam a sobrepor-se, não logo após as duas fendas, mas, agora, no segundo espelho semi-espelhado (***ESE₂***). E é esta sobreposição que cria uma distribuição da probabilidade de cada um dos fotões atingir as diferentes partes do detector, provocando o aparecimento da figura interferencial, após um grande número de impactos. Se onda ***θ*** existe, ou seja, se ela é uma onda física e não uma mera onda de probabilidade, de acordo com a posição realista de que parte o programa de investigação que estamos a expor, basta introduzir outra onda ***θ*** vazia, oriunda de uma segunda fonte monofotónica que, em princípio, não está correlacionada com a primeira, ou seja, que não é coerente com a primeira, e verificar o que acontece ao padrão interferencial. Pretende-se com este tipo de experiência verificar se a figura interferencial mantém a mesma forma antes e depois da introdução dessa onda ***θ***. Se observarmos uma alteração, esta experiência constituirá um forte argumento a favor da existência real das ondas ***θ***.

Para isso era necessário conceber uma fonte monofotónica que produzisse ondas ***θ***. Este problema foi resolvido da seguinte maneira[38]. Consideremos apenas a parte inicial do interferómetro de Mach-Zehnder representado na figura 4. Esta situação está representada na figura 5. Tal como acontece no interferómetro de Mach-Zehnder, temos uma fonte monofotónica e o primeiro espelho semi-espelhado separa a onda que constitui a partícula em duas ondas. Uma que acompanha o corpúsculo e que segue uma das vias, e outra, a onda ***θ*** que segue pela outra via. A porta ***P*** estará sempre fechada, não deixando passar nada, a não ser que o corpúsculo acompanhado pela parte da onda que atravessou o espelho semi-espelhado (***ESE***) atinja o detector ***D***. Nesse caso o detector determina a abertura da porta ***P*** deixando passar a parte da onda que tinha sido reflectida no espelho semi-espelhado (***ESE***). Se algum ser sairá da caixa, esse ser será uma onda ***θ*** vazia. Na situação em que o corpúsculo é reflectido no espelho semi-espelhado (***ESE***) nada sairá da caixa que encerra este gerador de ondas ***θ*** vazias. Aqui reside o ponto de divergência fundamental entre as duas interpretações do formalismo quântico.

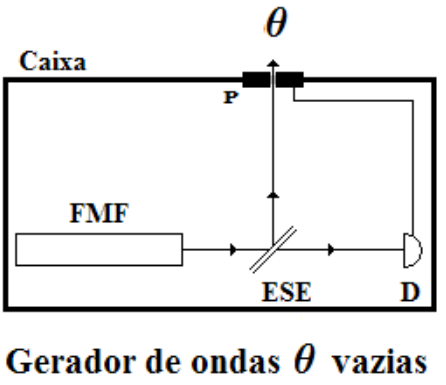


Figura 5

Para a interpretação bohriana do formalismo quântico ortodoxo, dado que a onda que representa uma partícula quântica é apenas uma onda matemática, ou seja, uma onda de probabilidade, e como a probabilidade de encontrar a partícula na via correspondente à sua reflexão no espelho semi-espelhado (***ESE***), após a detecção da

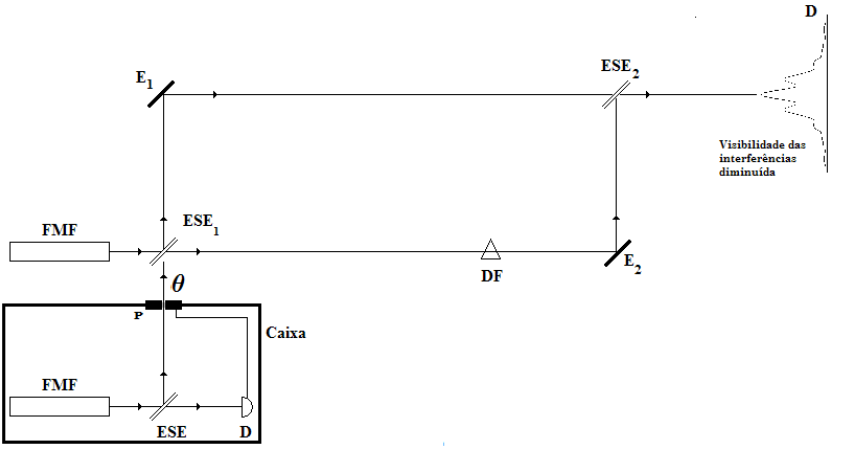
[38] J. R. Croca, in *Microphysical Reality and Quantum Formalism,* eds. A. van der Merwe, F. Selleri, G. Tarozzi, Kluwer, Dordrecht, 1988.

partícula no detector, é nula, nada seguirá por essa via e nenhum efeito poderá ser observado posteriormente. É isto que, como já referi, impõe o quinto postulado da mecânica quântica ortodoxa, ou seja, o postulado da redução instantânea do trem de ondas.

Para a interpretação brogliana do formalismo quântico não-linear e causal algo seguirá por essa via, porque é dessa forma que ela interpreta o que se passa numa experiência de interferências. O que segue por essa via é uma onda $\boldsymbol{\theta}$ vazia, uma onda não acompanhada pelo corpúsculo. Para esta interpretação bastará acoplar as figuras 4 e 5 e verificar se a figura interferencial obtida no caso representado na figura 4 é alterado. Os cálculos realizados mostram que se deve observar uma diminuição, facilmente observável, da visibilidade da figura interferencial em cerca de 30%. Isto corresponde a uma diminuição do contraste entre as zonas mais iluminadas e as zonas menos iluminadas da figura interferencial tal como era pretendido. Esta situação está representada na figura 6.

Para a interpretação da mecânica quântica ortodoxa nenhuma alteração se deveria verificar, pois quando a partícula foi detectada no detector existente no gerador de ondas $\boldsymbol{\theta}$ vazias, a onda de probabilidade, que corresponderia à alternativa de a partícula ter sido reflectida no espelho semi-espelhado reduzir-se-ia instantaneamente a zero, nada saindo do gerador de ondas $\boldsymbol{\theta}$ vazias. Como acabámos de ver, para um físico, e para a física, não é equivalente a adopção de uma posição instrumentalista ou de uma posição realista.

A adopção generalizadamente aceite de uma posição instrumentalista, pode ter provocado um enorme atraso no desenvolvimento da física. É bom que os filósofos da ciência conheçam profundamente a ciência e, além disso, participem na discussão dos seus fundamentos porque, de outra forma, correm o risco de cair numa neo-escolástica, como me parece terem caído muitos ao longo do século XX. É necessário voltar a olhar com atenção para o livro do universo como dizia Galileu.



Interferómetro de Mach-Zehnder com gerador de ondas θ vazias acoplado

Figura 6

Conclusão

Com este texto procurei mostrar um projecto de teoria que aparenta ter pernas para andar. Obedece às 5 primeiras condições. A existência das ondas $\boldsymbol{\theta}$ vazias, isto é, ondas desprovidas de corpúsculo (o contrário, de acordo com o compromisso ontológico proposto por Louis de Broglie e desenvolvido por J. R. Croca, nunca será verdadeiro, nunca poderá haver um corpúsculo sem onda) permite conceber resultados de experiências até aí impossíveis de conceber. As ondas $\boldsymbol{\theta}$ representam, neste compromisso ontológico, as características ondulatórias *reais* das partículas quânticas. Como acabámos de ver, neste novo quadro, a parte ondulatória de uma partícula desprovida de singularidade poderá alterar um padrão de interferências. Mas esta possibilidade é admitida como consequência da adopção de uma posição realista. Por sua vez, para a mecânica quântica ortodoxa, que adoptou uma posição instrumentalista das teorias físicas, a onda $\boldsymbol{\theta}$ pura e simplesmente não existe.

E quanto à última condição que devemos adoptar para considerar que um projecto de teoria é uma *boa*[39] teoria física? Alterará este, por enquanto, projecto de teoria a nossa capacidade de acção no mundo? As perspectivas são animadoras. Se o compromisso ontológico de que parte, no qual uma partícula quântica tem uma

[39] As teorias que obedecem aos 6 critérios enunciados.

estrutura complexa constituída por uma fortíssima concentração de energia (corpúsculo) e uma onda intimamente relacionada com esse corpúsculo, constituindo uma estrutura estável dentro daquilo a que David Bohm e Louis de Broglie chamaram o meio sub-quântico, não é inadmissível conceber processos de aproveitar a energia contida nesse, aparentemente inesgotável, reservatório. Esse reservatório de energia, no quadro do chamado "modelo standard" recebe o nome de "flutuações do vácuo", ou "energia do ponto zero", tem um carácter meramente virtual. Só num compromisso ontológico em que ele (meio sub-quântico) seja considerado como algo com conteúdo real faz sentido falar do aproveitamento da energia que ele contém, ou melhor, faz sentido procurar a forma de o conseguir. Somente no quadro desse compromisso é possível vislumbrar a forma de o conseguir. Isto porque esse compromisso permite entender os processos reais profundos que estão subjacentes aos fenómenos que observamos.

Por tudo o que acabei de dizer é-me impossível defender a equivalência de uma posição instrumentalista e de uma posição realista. É necessário reconhecer que a posição instrumentalista, dominante durante o desenvolvimento da física no século XX, levou-a a uma evidente estagnação. Essa atitude instrumentalista nasceu na mecânica quântica a partir de uma atitude céptica e, como seria inevitável, acabou por dar guarida a posições cépticas radicais. Pelo contrário, uma atitude realista como a que me esforcei por relatar neste trabalho põe-nos a um pequeno passo, tudo o indica, de importantes ganhos cognitivos.

Bibliografia

Livros

K. Popper, *Quantum Theory and the Schism in Physics*, Hutchinson, London, 1972.

J. R. Croca, *Towards a Nonlinear Quantum Physics*, World Scientific, 2003.

F. Selleri, *Lezioni di relatività da Einstein all'etere di Lorentz,* Progedit, Bari, 2003.

José R. Croca e Rui N. Moreira, *Diálogos sobre física quântica. Dos paradoxos à não-linearidade*, Esfera do Caos, Lisboa, 2007.

H. Høffding; É. Meyerson, *Correspondence entre Harald Høffding et Émile Meyerson*, ed. Frithiof Brandt, Hans Høffding et Jean Adigard des Gautries, (Einar Munskgaard), Copenhaga, 1939.

S. Rozental (ed.), *Niels Bohr. His life and work as seen by his friends and colleagues*, North-Holland Publishing Company – Amsterdam.

H. Høffding, *Esquisse d'une psychology fondée sur l'expérience*, Alcan, Paris, 1903.

H. Høffding, *La Relativité Philosophique* 1925, pp. 197-199. Esta tradução francesa engloba dois ensaios em dinamarquês de Høffding: *Totalitatet som* Kategori (Totalidade como categoria), Det kgl. D. Vid. S. Skrifter. 6. Række, hist. Filosofisk afd. 2. Copenhaga 1917, e *Relation som* Kategori (Relação como categoria), D. kgl. D. Vid. S. Filosofisk Meddelser, I, 3, Copenhaga, 1921.

J. Faye, *Niels Bohr: His Heritage and Legacy*, Kluwer Academic Publishers, Dordrecht, 1991.

Artigos

Robert L. Oldershaw, *American Journal of Physics*, Vol. 56 (12), December 1988.

Franco Selleri, comunicação privada.

Günter Nimitz, *Superluminal Signal Velocity and Causality,* Foundations of Physics, Vol. 34, No. 12, December 2004.

J. Croca, *Experimental Violation of Heisenberg's Uncertainty Relations*, talk at the 5th UK Conference on the *Conceptual and Philosophical Problems in Physics*, at Oxford, Sept. 1996.

Andrade e Silva, *Peut-on connaître simultanément et avec précision la position et la quantité de mouvement d'une particule?*, Portugaliæ Physica, 4 (1966) 257.

A.R. Silva, J.S. Ramos, J.R. Croca and R.N. Moreira, J.R. Croca, *NonLinear Schrödinger Equation, Burger's Equation and Superposition of Solutions*, in *Causality and Locality in Modern Physics and Astronomy*, eds. S. Jeffers et al. (Kluwer Academic Publishers, 1998).

J. and M. Andrade e Silva, C. R. Acad. Sci. (Paris) 290 (1980) 501.

F. Selleri, Ann.Found. L. de Broglie, 4 (1982) 45.

F. Selleri, Found. Phys. 12 (1983) 1087; 17 (1987) 739.

A. Garuccio, V. Rapizarda and J. P. Vigier, Phys. Lett. A 90 (1982) 17.

J. R. Croca, in *Microphysical Reality and Quantum Formalism*, eds. A. Van der Merwe, F. Selleri, G, Tarozzi, Kluwer, Dordrecht, 1988.

J. R. Croca, Ann. Found. L. de Broglie, 14 (1989) 323.

J. R. Croca, in Waves and Particles in Light and Matter, eds. Alwyn Van der Merwe and A. Garuccio, Plenum, 1994.

J. R. Croca, A. Garuccio, V. Lepori e R. N. Moreira, Found. of Phy. Lett. 3 (1990) 557.

J. R. Croca, R. Moreira and A. J. Rica da Silva, in *Causality and Locality in Modern Physics*, eds. G. Hunter et al., Kluwer Academic Publishers, 1998.

J. R. Croca, in *Microphysical Reality and Quantum Formalism,* eds. A. van der Merwe, F. Selleri, G. Tarozzi, Kluwer, Dordrecht, 1988.

Agradecimentos:

Agradeço ao meu amigo J. Croca a colaboração na revisão final do texto, bem como a colaboração na revisão das notas.

Este trabalho insere-se no projecto ***Problemas Filosóficos da Física Quântica*** financiado pela ***Fundação para a Ciência e a Tecnologia***.

CONSIDERACIONES SOBRE LA REFERENCIA DE LOS TÉRMINOS DE SUSTANCIA

Luis Fernández Moreno
Dpto. de Lógica y Filosofía de la Ciencia
Facultad de Filosofía
Universidad Complutense de Madrid

Resumen. En este trabajo voy a ocuparme de un tipo de términos que han estado en el centro del debate entre la teoría causal y la teoría descriptiva de la referencia. Estos términos son los términos de sustancias químicas, que suelen ser denominados de manera más abreviada "términos de sustancia", y esta terminología es la que emplearé en lo siguiente.
Los principales promotores de la teoría causal de los términos de sustancia son Kripke y Putnam, pero, puesto que de las teorías de estos autores la más elaborada es la de Putnam, en lo siguiente limitaré mis consideraciones a ésta. El objetivo de este trabajo consiste en examinar la teoría de la referencia de los términos de sustancia propuesta por Putnam y sus objeciones principales contra la teoría descriptiva de la referencia de este tipo de términos; en este marco esbozaremos una versión de la teoría descriptiva que se vea libre de dichas objeciones.

1. La teoría de la referencia de los términos de sustancia de Putnam

Putnam ha centrado sus consideraciones sobre los términos de sustancia principalmente en dos de estos términos, a saber, los términos "agua" y "oro", y ha formulado sus tesis acerca de la semántica de los términos de sustancia dentro de un contexto más amplio, como es el de los denominados *términos de género natural*, pues considera que los términos de sustancia son ejemplos prototípicos de términos de género natural. Más aún, el punto de partida de la teoría de Putnam acerca de los términos de género natural, tal como se presenta en (1988), lo constituyen justamente los términos de sustancia. Él afirma:

> "Una vez que se ha captado el aspecto fundamental en el caso de las sustancias, puede extenderse fácilmente a otros géneros naturales." (Putnam 1988, p. 34).

En su crítica a la teoría descriptiva de los términos de género natural Putnam supone que las descripciones a las que apelaría un seguidor de ese tipo de teoría expresarían sólo propiedades de carácter macroscópico y atribuye a la teoría descriptiva la tesis de que dichas propiedades constituirían el significado de los términos de género natural y determinarían su referencia.

Ahora bien, cabe distinguir entre la teoría descriptiva concebida como una teoría del significado o como una teoría de la referencia. Pues aunque los autores que sostengan una teoría descriptiva acerca del significado de un tipo de términos sostendrán de manera natural una teoría descriptiva acerca de su referencia, el sostener una teoría descriptiva acerca de la referencia de un tipo de términos no conlleva necesariamente sostener una teoría descriptiva acerca de su significado. En lo siguiente voy a asumir esta distinción, que extenderé a los términos de género natural, y limitaré mis consideraciones a las objeciones de Putnam contra la teoría descriptiva de la *referencia* de los términos de sustancia, los únicos tipos de términos de género natural que tomaré en consideración.

La tesis fundamental de Putnam acerca de la determinación de la referencia de los términos de género natural se basa en su concepción acerca de los géneros naturales. Esta concepción es la siguiente. En primer lugar, un género natural incluye un subconjunto de entidades a las que se atribuye el carácter de *miembros paradigmáticos* del género. En segundo lugar, la atribución de este rasgo a dichas entidades se debe a que comparten una serie de propiedades *macroscópicas* a las que cabe aludir como *propiedades distintivas normales* del género. En tercer lugar, suponemos que dichas entidades comparten *propiedades subyacentes*, que son causalmente responsables de dichas propiedades macroscópicas; en el caso de las porciones de agua la propiedad subyacente en cuestión es su composición química o, más precisamente, su estructura molecular, en el caso de las porciones de oro la propiedad subyacente es su número atómico o, si se prefiere, su estructura atómica. Dado este marco, la tesis fundamental de Putnam acerca de la determinación de la referencia de los términos de género natural es que su referencia viene determinada por la posesión de tales propiedades subyacentes. Puesto que son así propiedades subyacentes, y no las propiedades distintivas normales, las que determinan la referencia o extensión de un término de género natural, parece quedar abierta la posibilidad de que a esta extensión pertenezcan entidades carentes de

las propiedades distintivas normales y de que, en cambio, no pertenezcan a ella algunas entidades que sí las posean.

Llegados a este punto es pertinente atender a algunos detalles de la teoría de la referencia de Putnam acerca de los términos de género natural y, en especial, de los términos de sustancia.

La formulación clásica de la teoría de Putnam acerca de la referencia de los términos de género natural y, en concreto, de los términos de sustancia apareció formulada en su escrito, ya clásico, (1975). En ese artículo la teoría de la referencia de Putnam acerca de los términos de género natural se inserta justamente en una teoría acerca del significado de tales términos. Uno de los componentes del significado de los términos de género natural es el *estereotipo* asociado con el término, el cual incluye las propiedades que, de acuerdo con nuestras creencias de sentido común, caracterizan los miembros paradigmáticos del género, es decir, incluye las *propiedades distintivas normales*. De entre ellas las más centrales o más difícilmente revisables integran el *marcador semántico* del término, otro de los componentes del significado; así entre los marcadores semánticos del término "agua" estaría la propiedad de ser un líquido, mientras que el estereotipo de este término incluiría las propiedades de ser incoloro, transparente, insípido, de saciar la sed, etc. No obstante, como la distinción entre el estereotipo y el marcador semántico no es nítida, en lo siguiente hablaré simplemente del estereotipo del término.

Anteriormente ya hemos presentado la tesis fundamental de Putnam acerca de la determinación de la referencia o extensión de los términos de género natural, pero conviene señalar que Putnam hace hincapié en dos contribuciones involucradas en dicha determinación, a las que él denomina *la contribución del entorno* y *la contribución de la sociedad*. Por una parte, la extensión de un término de género natural depende de cómo es el mundo, pues viene determinada por propiedades subyacentes de miembros paradigmáticos del género pertenecientes a nuestro mundo. Por otra parte, la dilucidación de las propiedades subyacentes es objeto de la investigación científica y quienes la llevan a cabo o se sirvan de sus resultados – es decir, los *expertos* – tendrán un mejor conocimiento que el hablante medio acerca de las condiciones de pertenencia a la extensión de un término de género natural; hay a este respecto, en palabras de Putnam, una "división del trabajo lingüístico" (véase 1975, p. 228), de tal manera que el hablante medio está dispuesto a deferir en los expertos la

determinación de la referencia de los términos de género natural. En este sentido conviene señalar que hay una cierta inestabilidad en las exposiciones que el propio Putnam hace de su teoría, pues en ocasiones asimila su teoría a la teoría de Kripke , en la que la noción de bautismo inicial desempeña una importante función (véase Kripke 1980), pero en otras ocasiones Putnam da a entender claramente que los vínculos de carácter social entre el hablante medio y los expertos de su comunidad lingüística son más importantes para la referencia del término en su uso por dicho hablante que los vínculos de carácter histórico que enlazarían sus usos del término con la introducción inicial del mismo. De estas dos versiones de su teoría esta última parece ser la fundamental.

Por este motivo no es de extrañar que Putnam sustituya la cuestión de cómo tuvo lugar la introducción inicial de un término de sustancia como "agua" por la de cómo podemos enseñar a alguien el término "agua", fijando así la referencia de dicho término en su idiolecto. A este respecto él alude a dos procedimientos (véase 1975, pp. 229 y ss.). Uno de ellos consiste en una definición ostensiva; para ilustrarla supongamos que un hablante muestra a otro, quizás un aprendiz del castellano, un vaso que contiene agua y profiere la siguiente definición ostensiva: "Esto es agua" o, de manera más explícita, "Este líquido es agua".

Pero disponemos de otro procedimiento para fijar la referencia del término "agua" en el idiolecto de un hablante, a saber, mediante un tipo de *descripción* a la que Putnam denomina una *definición operacional*. Esta definición consiste del estereotipo asociado con el término "agua" en *nuestro mundo*, y su alcance es que una porción de líquido es una porción de agua si y sólo si está en la relación de identidad de líquido con los miembros paradigmáticos de la clase de entidades de nuestro mundo que satisfacen dicho estereotipo. Un aspecto común a ambas definiciones es que ambas involucran el empleo de *términos indéxicos*, como el término "este" – en la definición ostensiva – o como el término "nuestro" – en la cláusula "nuestro mundo" de la definición operacional –. Putnam alude a este rasgo, en el que se manifiesta la contribución del entorno a la referencia de los términos de género natural, afirmando que los términos de género natural poseen un *componente indéxico*, y alega que este rasgo conlleva que estos términos son *designadores rígidos*. Putnam afirma:

> "[U]na vez que hemos descubierto la naturaleza del agua, nada cuenta como un mundo posible en el que el agua no tenga esa naturaleza. Una vez que hemos descubierto que el agua (en el mundo actual) es H_2O, *nada cuenta como un mundo posible en el que el agua no sea* H_2O." (Putnam 1975, p. 233).

2. El argumento de la Tierra Gemela sobre el término "agua"

Es pertinente atender ahora a las principales objeciones de Putnam contra la teoría descriptiva de la referencia de los términos de sustancia. En (1975) Putnam presenta un tipo de experimentos mentales, a los que cabe aludir como "experimentos mentales de la Tierra Gemela" que tienen como finalidad hacer plausible la tesis de que los estados psicológicos, concebidos en sentido estrecho, no determinan la extensión de los términos de género natural. Ahora bien, puesto que Putnam asume que los rasgos involucrados en el estado psicológico, en sentido estrecho, en el que se encuentra un hablante que entiende un término de género natural se identificarían con las propiedades incluidas en el estereotipo asociado con dicho término, otra manera de entender la finalidad de dichos experimentos mentales, como haré en lo siguiente, es considerarlos dirigidos contra la tesis de que la referencia o extensión de los términos de género natural y, en concreto, de los términos de sustancia viene determinada mediante una descripción que expresaría las propiedades distintivas normales asociadas con dichos términos.

El primer experimento mental de la Tierra Gemela presentado por Putnam en (1975), y el único que tomaré en consideración, involucra el término "agua" (véase 1975, pp. 223-224). En este experimento mental cabe distinguir dos partes.

En la primera parte Putnam nos pide que imaginemos un planeta de *nuestra* galaxia que es muy semejante a la Tierra, al que vamos a denominar "Tierra Gemela", y en el que cada habitante de la Tierra tiene un duplicado, *completamente idéntico* a él. Aunque hay algunas pequeñas diferencias entre la Tierra y la Tierra Gemela, la única relevante para nuestros propósitos radica en que la composición química del líquido denominado "agua" en la Tierra Gemela no es H_2O, sino otra *muy diferente*, que abreviaremos como "XYZ". No obstante, *en condiciones normales* las sustancias cuya composición química es,

respectivamente, H_2O y XYZ son *macroscópicamente indistinguibles*. Imaginemos ahora que un habitante de la Tierra viaja a la Tierra Gemela; el supondrá, en primera instancia, a tenor de las propiedades distintivas normales de la sustancia denominada "agua" en la Tierra Gemela, que el término "agua" tiene el mismo significado (en el sentido de la misma *extensión*) en la Tierra y en la Tierra Gemela, pero él rechazará esta suposición tan pronto como descubra que la composición química de la sustancia presente en la Tierra Gemela no es H_2O, sino XYZ.

Putnam señala que la extensión del término "agua", tal como se emplea en la Tierra y en la Tierra Gemela, es *diferente*: la extensión del término "agua", tal como este término se emplea en la Tierra, consta de porciones de H_2O, mientras que la extensión del término "agua", tal como este término se emplea en la Tierra Gemela, consta de porciones de XYZ. Ahora bien, dada la contribución de *nuestro* entorno a la referencia de los términos de género natural, el líquido denominado "agua" en la Tierra Gemela no es agua.

Pasemos a la segunda parte del experimento mental. Putnam nos pide ahora que imaginemos que hemos hecho retroceder el tiempo hasta aproximadamente 1750, de tal manera que ni los habitantes de la Tierra ni los de la Tierra Gemela conocen todavía la composición química del líquido que denominan "agua". Tomemos ahora en consideración *un hablante típico* de la Tierra, $Óscar_1$, y su duplicado en la Tierra Gemela, $Óscar_2$; cada uno de ellos asociará con el término "agua" las *mismas* propiedades, a saber, las propiedades distintivas normales de la sustancia denominada "agua" en el planeta que habitan. No obstante, Putnam alega que la extensión del término "agua" en sus respectivos idiolectos es *diferente*. Aunque en 1750 todavía no lo sepan, las propiedades subyacentes del líquido al que denominan "agua" son diferentes.

Una vez presentado el experimento mental de la Tierra Gemela acerca del término "agua" conviene hacer algunas consideraciones críticas acerca del mismo. En primer lugar, los seres humanos estamos compuestos aproximadamente de un 60 % de agua; por lo tanto, $Oscar_2$ no es, a diferencia de lo que dice Putnam, "un duplicado" de $Oscar_1$. No obstante, el experimento mental puede reformularse con términos que designen sustancias que no formen parte de los seres humanos, por lo que la objeción en cuestión no es de peso. En segundo lugar, hay que señalar que, dada la *muy diferente* composición química de la sustancia denominada "agua" en la Tierra y en la Tierra Gemela, es cuestionable suponer, como hace Putnam,

que esas sustancias son *macroscópicamente indistinguibles*. Pero, siendo esto así, la plausibilidad del experimento mental se desvanece, pues en 1750 $Oscar_1$ y $Oscar_2$ asociarían con el término "agua" distintas propiedades macroscópicas y si se alega, como en ocasiones hace Putnam, que la indiscernibilidad macroscópica de esas sustancias sólo se da en *condiciones normales*, cabe replicar que, puesto que se supone que $Oscar_1$ y $Óscar_2$ son hablantes típicos respectivamente de la Tierra y de la Tierra Gemela, sus interacciones con la sustancia que denominan "agua" involucrarán el estado de la sustancia respectiva sólo en condiciones normales; obviamente cabe suponer que en 1950 $Oscar_1$ y $Oscar_2$ asociarán con el término "agua" distintas propiedades subyacentes, a saber, las de ser respectivamente H_2O y XYZ.

Por último, conviene señalar que variaciones en el experimento mental de la Tierra Gemela llevan a rechazar una de las consecuencias que Putnam extrae de dicho experimento mental, según la cual el líquido denominado "agua" en la Tierra Gemela no es agua. Así supongamos que algunos habitantes de la Tierra hubiesen sido trasladados por extraterrestres a la Tierra Gemela en 1750 y hubiesen comenzado a vivir allí. Ellos asociarían con el líquido denominado "agua" en la Tierra Gemela las mismas propiedades que habían asociado con el líquido denominado "agua" en la Tierra y lo utilizarían para los mismos menesteres, por lo que considerarían que se trata del *mismo* líquido – supongamos, como hace Putnam, que en condiciones normales dichos líquidos son macroscópicamente indistinguibles –. Pasado cierto tiempo y con el avance de la ciencia tanto en la Tierra como en la Tierra Gemela se descubre que la composición química del líquido denominado "agua" en la Tierra Gemela es XYZ. Supongamos adicionalmente que el avance científico concernió también a los medios de comunicación con sus familiares terrícolas, que les informan de que se había descubierto que la composición química del líquido denominado "agua" en la Tierra es H_20. En esta situación estos antiguos habitantes de la Tierra que se habían trasladado a la Tierra Gemela considerarían que hay dos tipos de agua, que se diferencian por su composición química. Y esto es lo que comunicarían a sus familiares habitantes de la Tierra. Por tanto, variaciones en el experimento mental de la Tierra Gemela permiten extraer una conclusión diferente de la de Putnam: hay *dos* tipos de agua. Pero para encontrarnos en una situación similar no hace falta recurrir a experimentos mentales. El propio Putnam pone el ejemplo del término "jade", que se aplica a dos

minerales, jadeita y nefrita, que poseen cualidades táctiles un tanto similares, aunque su microestructura es diferente. Obviamente, la diferencia es que la jadeita y la nefrita son minerales que se encuentran en la Tierra, mientras que el líquido cuya composición es XYZ se encuentra en la Tierra Gemela, pero éste es un planeta de *nuestra* galaxia. Por este motivo, no es de extrañar que haya autores que sostengan que ante el experimento mental de Putnam quepa replicar que hay dos tipos de agua, que sólo se diferenciarían por su microestructura.

Ahora bien, incluso aceptando los argumentos de Putnam y sus presupuestos, a saber, la concepción de Putnam acerca de los géneros naturales y, por tanto, acerca de la determinación de la referencia de los términos de género natural, nada hay en todo ello que constituya una objeción decisiva contra la teoría descriptiva de la referencia.

3. La teoría descriptiva de la referencia de los términos de sustancia

A este respecto conviene hacer al menos tres tipos de consideraciones. En primer lugar, no hay razón por la que la teoría descriptiva haya de asumir que la referencia de un término de género natural y, por tanto, de un término de sustancia venga determinada sólo mediante descripciones que expresan las propiedades distintivas normales asociadas con el término. De hecho, los seguidores actuales de la teoría descriptiva, como F. Jackson, apelan a una noción muy amplia de propiedad, y no como suponen los críticos de la teoría descriptiva, a propiedades puramente cualitativas; así Jackson afirma:

> "El término 'propiedad' [...] es simplemente una abreviación de una manera en la que las cosas podrían ser en un sentido amplio, que incluye maneras relacionales y disposicionales en las que las cosas podrían ser." (Jackson 1998, p. 202).

En segundo lugar, la teoría descriptiva puede aceptar las definiciones ostensivas y operacionales de Putnam. Aunque ambas definiciones contienen términos indéxicos, los principales promotores contemporáneos de la teoría descriptiva de la referencia, como J. Searle, incluyen entre las descripciones identificadoras que determinarían la referencia de los nombres propios, y lo mismo se

aplicaría a los términos de sustancia, tanto las descripciones indéxicas, involucradas en la definición ostensiva de los términos de sustancia, como las descripciones mixtas, involucradas en las definiciones operacionales de los términos de sustancia. Así afirma Searle:

> Por "descripción identificadora" se entiende una expresión de cualquiera de los tres tipos siguientes: "presentación demostrativa, descripción única […] [e] identificación mixta demostrativa y descriptiva." (Searle 1969, p. 86).

Ahora bien, puesto que la referencia de los términos indéxicos depende del contexto de proferencia, y a este respecto la Tierra y la Tierra Gemela constituyen contextos diferentes, las definiciones ostensivas o las definiciones operacionales mediante las cuales se hubiese enseñado a $Oscar_1$ y $Oscar_2$ el término "agua" y se hubiese fijado la referencia de ese término en su idiolectos, determinarían, como pretende Putnam, sustancias diferentes, a saber, respectivamente, H_2O y XYZ.

En tercer lugar, hay otro tipo de descripciones que pueden ser importantes para la determinación de la referencia de los términos de sustancia, a saber, las descripciones que expresan las propiedades subyacentes de una sustancia que un experto puede proporcionar, si bien éstas son revisables. Obviamente, el hablante medio no conoce, por regla general, tales descripciones, si bien, como señala Putnam, dicho hablante estará dispuesto a deferir la extensión de los términos de sustancia tal como los utiliza en su uso por parte de los expertos de su comunidad lingüística. Así, p.ej., aunque el hablante medio no conoce el número atómico del oro, si dicho hablante posee las intenciones deferenciales que Putnam le atribuye en relación con los términos de género natural, dicho hablante asociará *implícitamente* con el término "oro" una descripción como la siguiente: la sustancia a la que se refieren los expertos de mi comunidad lingüística acerca del oro cuando profieren la palabra "oro". Estas descripciones, a las que podemos denominar "parásitas", sólo serán identificadoras en la medida en que remitan a otras descripciones asociadas por los expertos que, por sí mismas, sean identificadoras, como lo es la descripción "la sustancia cuyo número atómico es 79". Por tanto, entre las propiedades que el hablante medio asociará con un término de

sustancia no se encontrarán sólo aquellas correspondientes a las propiedades distintivas normales de una sustancia.

De esta manera si se acepta la tesis de Putnam de que el término "agua" tal como es usado en la Tierra y en la Tierra Gemela tiene distinta referencia, la teoría descriptiva puede aceptar igualmente dicha tesis, incorporando tanto la contribución del entorno a la determinación de la referencia como la contribución de la sociedad. A este último respecto conviene señalar que Putnam afirma:

> "[La palabra] 'agua' […] no la exhibió [e.d., la división del trabajo lingüístico] […] antes del surgimiento de la química." (Putnam 1975, p. 228).

Por tanto, según Putnam, a partir de esa época – es decir, hacia finales del siglo XVIII – habría expertos acerca del agua y sus descripciones pueden tener relevancia para la determinación de la referencia del término "agua".

4. La referencia de los términos de sustancia en su uso cotidiano: las sustancias impuras

Muchos de los términos de sustancia que han estado en uso antes del surgimiento de la ciencia o, en la terminología de Putnam, antes de haber pasado a estar sujetos a la división del trabajo lingüístico, como el término "agua", tienen un uso en el lenguaje científico, pero también un uso en el lenguaje cotidiano y los criterios de aplicación correspondientes a esos dos tipos de usos no son idénticos. La propuesta fundamental de Putnam acerca de los términos de sustancia en su uso cotidiano es que estos términos se refieren a entidades que, además de poseer la estructura subyacente de la sustancia en cuestión, que es la más determinante de su referencia, contienen otros ingredientes o, como él acostumbra a decir, *impurezas*. Por tanto, los criterios de aplicación de los términos de sustancia en su uso cotidiano poseen un carácter derivado con respecto a los criterios correspondientes a su uso en la ciencia. Esta tesis no es de extrañar, pues Putnam formula su teoría acerca de los términos de sustancia en el marco de una teoría de los términos de género natural, pero cabe cuestionar que la extensión de un

término de sustancia, como "agua", en su uso cotidiano concuerde con la concepción de Putnam acerca de los géneros naturales.

Conviene señalar que incluso la concepción de los géneros naturales, tal como suele formularla Putnam, es un tanto simplificada. Putnam identifica la microestructura del agua con H_2O. Ahora bien, esta identificación encierra una simplificación, por cuanto tanto el hidrógeno como el oxígeno tienen diferentes tipos de isótopos, por lo que, en sentido estricto, el término "agua", en su uso en ciencia, "el agua químicamente pura", se refiere a mezclas de distintas sustancias (véase, por ejemplo, Zemach 1976). Sobre esta puntualización volveremos en alguna ocasión, aunque, por regla general y para abreviar, nos serviremos, como hasta ahora, de la simplificación mencionada.

Ahora bien, las sustancias como el agua no se encuentran de manera pura en la naturaleza y aplicamos el término "agua" no sólo a porciones de sustancia – y, especialmente, a porciones de líquido – que están compuestas exclusivamente de H_2O, sino también a otras que están compuestas de H_2O y de otros ingredientes; piénsese, por ejemplo, en una porción de líquido tomada del Mediterráneo, o de un río, o del grifo. Pero hay asimismo porciones de líquido que contienen H_2O y otros ingredientes que no pertenecen a la extensión del término "agua"; piénsese, por ejemplo, en el contenido de una taza de té o de café. En estos ejemplos se observa que la apelación exclusiva a propiedades subyacentes, como la propiedad de estar compuesto de H_2O, no permite determinar por sí sola la extensión del término "agua" en su uso cotidiano, es decir, en su uso más frecuente en un lenguaje natural. Además de apelar a esta propiedad habrán de tomarse en consideración otros tipos de propiedades. Y a este respecto la apelación a *propiedades distintivas normales* es sumamente relevante y es la que, por ejemplo, nos lleva a excluir de la extensión del término "agua" el contenido de una taza de té e incluir en su extensión una porción de líquido tomada del grifo; consideraciones de este tipo llevan a sostener que, por regla general, sólo en combinación con propiedades distintivas normales del agua, la propiedad de estar compuesta de H_2O puede determinar la extensión del término "agua".

La extensión del término "agua", en su uso cotidiano, es así imprecisa, pues para su determinación hay que tomar en consideración no sólo propiedades subyacentes, sino también propiedades distintivas normales, y la importancia concedida a unas u otras dependerá de

factores como el contexto y los intereses de los hablantes, factores que Putnam ya había mencionado, aunque muy de pasada en (1975) (véase 1975, pp. 239 s.), y ha subrayado en obras posteriores. Ésta sí es *la extensión predominante* del término "agua", al menos si entendemos por "la extensión predominante" de un término la que corresponde al uso más frecuente del mismo.

A este respecto conviene señalar que en ocasiones Putnam hace afirmaciones difícilmente sostenibles; así afirma:

> "Bajo el supuesto de que el agua normal es en efecto una sustancia *pura*, no pretendemos que […] [el término 'agua' se aplique] a algo que conste en alto grado (digamos, un veinte por ciento o más) de otra sustancia." (Putnam 1988, p. 31; mi cursiva).

Conviene señalar que en este texto Putnam está entendiendo por "sustancia pura" una sustancia que consta exclusivamente de moléculas de H_2O. Ahora bien, ni el agua *normal* es una sustancia pura ni hay razones para fijar en un veinte por ciento el límite máximo de otros ingredientes distintos de moléculas de H_2O a partir del cual a una porción de líquido no se le aplicaría correctamente el término "agua".

La apelación a porcentajes se encuentra también en otros textos de Putnam, como en el siguiente:

> "Por ejemplo, si el científico le dice al lego que el 50 por ciento del líquido en el vaso que acaba de beber es, de hecho, una sustancia química (inofensiva) que no figura como un constituyente del agua normal, el lego *no* dirá [que es agua]." (Putnam 1990, p. 326, n. 12).

Poniéndonos en el caso del hablante lego, la verdad es que en un caso como éste no se sabe muy bien qué decir; si el científico nos asegura que el otro 50 por ciento del líquido constaba de moléculas de H_2O, y si además el líquido poseía las propiedades distintivas normales del agua, quizás sí diríamos que era agua.

En cualquier caso, la tesis más plausible acerca del término "agua", en su uso cotidiano, es que su extensión viene determinada tanto por propiedades subyacentes como por las propiedades distintivas normales del agua.

No obstante, alguien podría replicar que el término "agua" es un caso peculiar, por lo que puede ser conveniente atender al término "oro". Pero si tomamos en consideración el término "oro" en su uso predominante en el lenguaje natural, la referencia del término "oro" no viene determinada exclusivamente por la propiedad de ser el elemento con el número atómico 79 – en sentido estricto, una mezcla de isótopos de dicho elemento –. Esto sólo sería así si por "oro" se entiende el oro químicamente puro. Pero el oro empleado en joyería no es oro químicamente puro, pues es demasiado blando para hacer joyas. Las joyas de oro son aleaciones de oro químicamente puro y de otros metales, como cobre, plata, etc. Por tanto, tal como los joyeros emplean el término "oro", su extensión no viene determinada exclusivamente mediante la propiedad subyacente mencionada y puesto que el hablante medio está dispuesto a delegar en ellos la referencia del término "oro" tal como lo usa, la extensión del término "oro" en lo que podríamos denominar su "uso cotidiano" o, en cualquier caso, en su uso más frecuente en el lenguaje natural, tampoco viene determinada exclusivamente mediante dicha propiedad.

De esta manera ni la extensión del término "agua" ni la extensión del término "oro", en sus usos más frecuentes en el lenguaje natural, están determinados mediante *una* propiedad subyacente y esas extensiones son un tanto imprecisas, pues en su determinación intervienen tanto propiedades subyacentes como propiedades distintivas normales.

Más aún Putnam ha de reconocer que en muchos casos estas últimas son las decisivas. Hay una sección de (1975) titulada "Otros sentidos", que aparece como una especie de paréntesis en el escrito, donde Putnam hace dicha concesión. Así afirma:

> "[S]i hay *una* estructura oculta, entonces generalmente determina en qué consiste ser un miembro del género natural, no sólo en el mundo actual, sino en todos los mundos posibles. En otras palabras, determina lo que cabe y no cabe suponer contrafácticamente acerca del género natural […]. Pero el agua local, o lo que sea, puede tener dos o más estructuras ocultas – o tantas – que la 'estructura oculta' se torna irrelevante y las características superficiales se convierten en las decisivas." (Putnam 1975, p. 241).

Ahora bien, la posibilidad mencionada en la última parte de este texto es la que corresponde a nuestra situación actual, especialmente si tomamos en consideración el caso de los isótopos de hidrógeno y oxígeno. No obstante, la conclusión a extraer no es que las propiedades subyacentes se tornen irrelevantes, sino que *también* las propiedades macroscópicas o propiedades distintivas normales desempeñan un papel importante en la determinación de la referencia de los términos de sustancia, y especialmente en su uso en el lenguaje cotidiano.

Por tanto, cabe concluir que los términos de sustancia, o al menos aquellos que han estado en uso antes del surgimiento de la ciencia moderna, tienen, cuando menos, dos tipos diferentes de usos, y aunque éstos no sean completamente independientes, son diferentes: su uso en la ciencia y su uso cotidiano. La teoría de Putnam puede dar cuenta, con algunas simplificaciones, del primer tipo de uso, pero le resulta más difícil dar cuenta del segundo, a no ser que apele a factores como las propiedades distintivas normales, a los que él tendía sólo a otorgar una importancia muy secundaria y cuya relevancia para la determinación de la referencia limitaba en principio a la *teoría descriptiva*.

Referencias bibliográficas

Jackson, F. (1998). "Reference and description revisited", en J. E. Tomberlin (ed.), *Language, Mind, and Ontology* (*Philosophical Perspectives*, vol. 12), Blackwell, Oxford, pp. 201-218.

Kripke, S. (1980). *Naming and Necessity*. Oxford: Blackwell. Reimp. revisada y con prefacio añadido de "Naming and necessity", en D. Davidson y G. Harman (eds.), *Semantics of Natural Language*, Reidel, Dordrecht, 1972.

Putnam, H. (1975). "The meaning of 'meaning' ", en H. Putnam *Mind, Language and Reality*. (*Philosophical Papers*, vol. 2). Cambridge, Cambridge University Press, pp. 215-271.

_____. (1988). *Representation and Reality*. Cambridge: MIT Press.

_____. (1990). *Realism with a Human Face*. (Edición e introducción a cargo de J. Conant). Cambridge, Mass.: Harvard University Press.

Searle, J. (1969). *Speech Acts*. *An Essay in the Philosophy of Language*. Cambridge: Cambridge University Press.

Zemach, E. M. (1976). "Putnam's theory of the reference of substance terms". *Journal of Philosophy* 73, pp. 116-127.

MÉTODOS FORMALES EN COMPUTACIÓN BIO–INSPIRADA

Mario J. Pérez-Jiménez
Grupo de Investigación en Computación Natural
Dpto. de Ciencias de la Computación e Inteligencia Artificial
Universidad de Sevilla

Resumen. La Naturaleza viva es una fuente de inspiración para el diseño de modelos de computación susceptibles de ser implementados en dispositivos con soporte biológico. Los trabajos pioneros de L. Adleman (1994) y Gh. Păun (1998) sentaron las bases de dos modelos no convencionales: el primero inspirado en la manipulación de moléculas de ADN y el segundo inspirado en la estructura y funcionamiento de las células de los organismos vivos.
En este trabajo se describen los elementos básicos de dichos paradigmas computacionales y se analiza la verificación formal de los procedimientos mecánicos en dichos modelos que resuelven ciertos problemas. Para ello se asocia un sistema formal a cada modelo de tal manera que el proceso de verificación equivale a probar la adecuación y completitud del sistema formal correspondiente.

1. Introducción

Una constante de la humanidad a lo largo de toda la historia ha sido el planteamiento de problemas cuya resolución tiene como objetivo final la mejora de la calidad de vida. El estudio de los problemas planteados ha conducido, de manera natural, a la búsqueda de procedimientos sistemáticos que permitieran su resolución sin necesidad de tener un conocimiento profundo de los mismos.

A lo largo de este artículo, el término problema se entiende como sinónimo de *problema abstracto* que es una colección de problemas concretos, denominadas *instancias* o datos de entrada del problema abstracto. Por ejemplo, un problema abstracto sería: *dados dos números naturales, determinar su máximo común divisor*; una instancia de ese problema abstracto sería, por ejemplo: *hallar el máximo común divisor de los números 1578 y 3651696*.

Es interesante observar que dado un problema abstracto, las distintas instancias del mismo tienen asociadas, de manera natural e intuitiva, un concepto de *tamaño*: Por ejemplo, las instancias *hallar el máximo común divisor de los números 45 y 36165* y *hallar el máximo común divisor de los números 1578 y 3651696* tienen diferente tamaño y, es lógico pensar que el segundo tiene un tamaño *mayor* que el primero.

En la vida real aparecen numerosos problemas (concretos) que interesan resolver. Una metodología general consistiría en lo siguiente:

- Abstraer del problema concreto los hechos que se consideren *más relevantes*.
- A partir de ese proceso, diseñar un problema abstracto que pueda ser un *buen modelo* del problema concreto.

De esta manera es posible que muchos y muy diferentes problemas sean modelados por el mismo problema abstracto.

Por ejemplo, una determinada empresa *A* tiene una flota de 10 camiones cada uno de los cuales tiene una determinada capacidad y recibe cada día el encargo de otra empresa *B* para que le distribuya entre sus clientes una serie de objetos (electrodomésticos, artículos de alimentación, artículos de regalos, etc.). Cada uno de esos objetos ocupan un determinado volumen y tiene asignado dos parámetros: la dirección en donde hay que entregarla y una cantidad económica que la empresa *B* le da a la empresa *A* por ese reparto. Cada día, la empresa *A* se enfrenta a un problema concreto: realizar el reparto a fin de que los beneficios obtenidos sean máximos. Pudiera suceder que la empresa no reparta todos los objetos durante un día.

Parece claro que todos esos problemas concretos que tiene la empresa *A* son, realmente, instancias de un problema abstracto: *Dados n camiones con capacidades* $c_1, \ldots c_n$ *y dados p objetos* $o_1, o_2, \ldots, o_p$, *cada uno de los cuales tienen asignados una terna de números* $(v_1, d_1, b_1), \ldots, (v_p, d_p, b_p)$ hallar una distribución de los objetos entre los camiones y unas rutas para cada camión de tal manera que la suma de los beneficios obtenidos por los objetos repartidos ese día sea máxima.

Para la empresa *A* será importante disponer de una solución del problema abstracto asociado y, entonces, cada día basta considerar la

instancia correspondiente y "ejecutarla" de acuerdo con la solución general. Entonces sería muy interesante que la solución del problema general fuera de tipo *mecánico* a fin de que la obtención de la solución del problema concreto de cada día se redujera a ejecutar (sobre una máquina) la solución general con los correspondientes datos de entrada del problema concreto.

Una buena definición formal de problema abstracto, desde el punto de vista computacional, sería la siguiente: es una terna $X = (E_X, S_X, f_X)$ en donde E_X y S_X son conjuntos de palabras (*lenguajes*) sobre un cierto alfabeto finito, y f_X es una aplicación de E_X en S_X. Es decir, un problema consta de un conjunto de datos de entrada, E_X, un conjunto de datos de salida, S_X, y una función, f_X, que asigna a cada dato de entrada del problema, una salida correcta del mismo.

Un *problema de decisión* está caracterizado informalmente por la condición de que sólo admite dos salidas posibles: *Sí* o *No*. De ahí que un problema de decisión, *X*, se pueda identificar con un par (E_X, θ_X), en donde E_X es el conjunto de datos de entrada y θ_X es un predicado sobre E_X que a cada dato de entrada del problema le asigna el valor verdadero (*Sí*) o falso (*No*).

Así pues, para resolver un determinado un problema concreto de la vida real hay que proceder como sigue:

- Diseñar un problema abstracto que modele ese problema concreto.
- Diseñar una solución (a ser posible, mecánica) de ese problema abstracto.
- Ejecutar la solución del problema abstracto suministrándole los datos específicos de la instancia del problema concreto de la vida real.

No obstante, pudiera suceder que a la hora de ejecutar ciertas instancias la máquina necesitara demasiado tiempo para realizar los cálculos correspondientes, o bien, necesitara usar una *memoria* que excediera las capacidades de la máquina.

La aparición de los primeros métodos formales de razonamiento (posiblemente en Mesopotamia, y mejorados sustancialmente en la civilización griega) proporcionó una herramienta interesante para poder expresar de forma algo más precisa el concepto intuitivo de *procedimiento sistemático, procedimiento mecánico* o *algoritmo*.

Un modelo de computación consiste, básicamente, en formalizar el concepto de *procedimiento mecánico*. Para ello, existen diferentes metodologías:

- Definir en un marco formal la sintaxis de los procedimientos mecánicos (que permite distinguir *físicamente* un tal procedimiento) así como la semántica (es decir, cómo se ejecutan esos procedimientos cuando se le suministra un dato de entrada). A partir de esta formalización, se introducen los conceptos de *función computable* y *problema resoluble* en el modelo. Este tipo de modelos de computación se denominan *orientados a programas*. Ejemplos de modelos de este tipo son todos aquellos que consideran los procedimientos mecánicos como un programa escrito en un lenguaje de programación, a partir de unas determinadas instrucciones elementales y una semántica asociada a ellas.
- Considerar una clase de funciones distinguidas como las *funciones computables* del modelo. A partir de las cuales se introducen los conceptos de *problema resoluble* y *procedimiento mecánico*. Este tipo de modelos de computación se denominan *orientados a funciones* y son ejemplos significativos el modelo de las funciones recursivas introducido por K. Gödel entre 1931 y 1933, y el de las funciones λ–calculables introducidos por A. Church y S. Kleene en 1931.
- Definir el concepto de máquina del modelo capaz de ejecutar procedimientos mecánicos que se pueden expresar a modo y manera de programas (por ejemplo, el modelo de las máquinas URM, introducido por Shepderdson y Sturgis en 1965) o bien mediante procesos implícitos, como las funciones de transición (por ejemplo, el modelo de las máquinas de Turing, introducido por A. Turing en 1936). Este tipo de modelos de computación se denominan *orientados a máquinas*.

Todo modelo de computación consta de: (a) una *estructura de datos* que serán los objetos susceptibles de ser manipulados en el modelo; (b) una *sintaxis* que precisa las reglas gramaticales para la construcción de los *procedimientos mecánicos del modelo*; y (c) una *semántica* que define la forma en que los procedimientos mecánicos operan sobre los datos de la estructura.

A finales de la década de los cuarenta comienzan a diseñarse los primeros ordenadores de propósito general y en la década de los cincuenta se desarrollan los primeros *lenguajes de programación*, *traductores de lenguajes* y *sistemas operativos*. La potencia de los ordenadores en esa época estaba muy limitada por la excesiva lentitud de los procesadores y la escasa memoria de la que disponían para almacenar la información.

No obstante, la aparición de estas primeras máquinas generales de cálculo provoca un cambio de orientación en ciertas investigaciones. Hasta entonces, un problema central consistía en la clasificación de los problemas abstractos según fueran o no resolubles mecánicamente en un cierto modelo de computación *universal*. A partir de ahora, la comunidad científica empieza a preguntarse qué problemas abstractos pueden ser resueltos por máquinas reales (ya construidas) y para instancias del problema *suficientemente grande*. Desde luego, estas ideas informales habría que precisarlas y así nació la *Teoría de la Complejidad Computacional* en la década de los sesenta, por parte de J. Hartmanis, P.M. Lewis y R.E. Stearns [23][27]. A partir de este momento se impulsa el análisis de los recursos computacionales que un algoritmo necesita para su ejecución. Dicho estudio requiere el uso de técnicas matemáticas (inducción, ecuaciones de recurrencia, notaciones asintóticas, manipulación de sumas finitas, etc.) que, además, permitan un análisis comparativo de distintas soluciones algorítmicas de un mismo problema.

En la década de los cincuenta, el premio nobel R.P. Feynman [15] describe los ordenadores *sub–microscópicos* e introduce el concepto teórico de *computación a nivel molecular*, postulándolo como una innovación necesaria y revolucionaria en la carrera por la miniaturización. Las ideas de Feynman adquieren una especial relevancia a partir de 1983, cuando R. Churchhouse establece las limitaciones fìsicas de la velocidad de cálculo de un procesador convencional. Con ello, serìa imposible resolver problemas

importantes de la vida real mediante el uso de máquinas que tuvieran un soporte electrónico.

Por tanto parece necesario la búsqueda de vías alternativas para considerar nuevos modelos de computación que pudieran ser implementados en soportes no electrónicos.

2. Modelos de Computación No–convencionales Bio–Inspirados

La *Computación Natural* bio–inspirada surge como una posible alternativa a la computación que podríamos denominar convencional, en la búsqueda de nuevos paradigmas que puedan proporcionar una solución *efectiva* a problemas de la vida real que jamás podrán ser resueltos mediante el uso de máquinas electrónicas.

Actualmente, dentro del concepto de *Computación Natural* se engloba un conjunto de modelos que tienen como característica común la simulación del modo en que la naturaleza actúa/opera/calcula sobre la materia. El objetivo central de la Computación Natural es la formalización, implementación y simulación de procedimientos dinámicos que se dan en la Naturaleza y que son susceptibles de ser interpretados como procedimientos de cálculo. Esta disciplina surge de la investigación de modelos matemáticos y de las herramientas tecnológicas necesarias para la implementación de la computación bio–inspirada. Entre las áreas que se enmarcan dentro de la Computación Natural, destacamos las siguientes:

1. Los *algoritmos genéticos* (o más en general, la *computación evolutiva*), introducidos por J. Holland [25] en 1975, que hacen uso de algunas operaciones inspiradas en la evolución y en la selección natural a fin de encontrar una *buena* solución a partir de una gran cantidad de posibles soluciones candidatas.

2. Las *redes neuronales artificiales*, introducidas por W.S. McCulloch y W. Pitts [30] en 1943, que están inspiradas en las interconexiones y en el funcionamiento de las neuronas en el cerebro.

3. La *computación molecular*, cuyo objetivo consiste en usar moléculas orgánicas (ADN, ARN, proteínas, etc.) como hardware biológico que permita realizar computaciones. Esta

disciplina nace a finales de 1994 con los trabajos de L. Adleman [1], si bien tiene precedentes en un trabajo de T. Head [24] en el que formula el *sistema splicing* (modelo teórico de procesamiento de moléculas de ADN con la participación de enzimas).

4. La *Computación Celular con Membranas*, introducida por Gh. Păun [32] en 1998, que está inspirada en la estructura y el funcionamiento de las células de los organismos vivos, en cuanto a su capacidad para procesar y generar información.

Los algoritmos genéticos y las redes neuronales artificiales han sido implementadas a través de programas en ordenadores electrónicos convencionales. La computación molecular basada en ADN ha sido implementada en medios bioquímicos (el experimento de L. Adleman permitió resolver una instancia concreta del problema del camino hamiltoniano, en su versión dirigida y con dos nodos distinguidos, a través de la manipulación de moléculas de ADN en el laboratorio). En cambio, la Computación Celular con Membranas aún no ha sido implementada ni en medios electrónicos ni bioquímicos.

3. Modelos de Computación Molecular

Los modelos de computación molecular son modelos *orientados a programas*. Las operaciones elementales o básicas estarán inspiradas en operaciones bioquìmicas con moléculas de ADN (que se denominarán *operaciones moleculares*), si bien usaremos, además, operaciones (como bucles y asignaciones) cuya finalidad es determinar el orden de ejecución de las operaciones moleculares.

Los modelos moleculares operan sobre unos unos objetos que denominamos *tubos* (asociados a un alfabeto arbitrario prefijado) y que son abstracciones de los tubos de ensayo. Cada tubo estará compuesto de un cierto *multiconjunto* (conjunto cuyos elementos pueden aparecer repetidas veces) de cadenas sobre un alfabeto, cada una de las cuales codificará una molécula de ADN.

Todo modelo molecular quedará determinado por la especificación de las operaciones moleculares elementales. La *función semántica* del modelo se define de manera natural para las instrucciones moleculares (abstracción del resultado correspondiente

de la operación molecular que representa). Un *programa molecular* es una una sucesión finita de instrucciones del modelo que admite como dato de entrada un tubo y, además, el dato de salida puede ser `sí`, `no`, o bien un elemento (que representa una molécula) del tubo final.

3.1. Modelo restringido de Adleman

En esta sección vamos a presentar el primer modelo molecular, denominado *restringido*, introducido por L. Adleman [2] en 1995. Para ello, debemos especificar las operaciones moleculares básicas de dicho modelo.

Se trata de un modelo de computación basado en procedimientos de *filtrado*; es decir, las computaciones en dicho modelo parten de un tubo inicial que contiene todas las posibles soluciones del problema y, mediante un determinado criterio, se procede a realizar sucesivos filtrados hasta obtener un tubo de salida que debe contener todas las soluciones correctas del mismo, y sólo esas. Además, en dicho modelo las computaciones no alteran la estructura interna de las moléculas que aparecen en los tubos; es decir, la información que codifica cada molécula es invariable a lo largo de todo el proceso. Por ello, se suele decir que estos modelos de computación molecular carecen de memoria.

A la hora de trabajar en estos modelos hay que distinguir claramente una *fase de inicialización o precomputación* y una *fase de ejecución*. En la primera se construye un tubo de ensayo inicial que debe contener cadenas que codifican todas las posibles soluciones del problema; en la segunda se van filtrando del tubo inicial las moléculas que codifican soluciones correctas del problema (y sólo esas).

Definición 1 *Un* **agregado** *sobre un alfabeto Σ es un multiconjunto finito de elementos de Σ. Un* **tubo** *sobre un alfabeto Σ es un multiconjunto finito de agregados sobre Σ.*

Las instrucciones moleculares básicas del modelo restringido de L. Adleman son las siguientes:

• **Extraer(*T,s*)**: esta operación admite como entrada un tubo T y un *sìmbolo* $s \in \Sigma$, y devuelve como salida dos tubos:

$$+(T,s)=\{\sigma \in T: s \in \sigma\} \quad \text{y} \quad -(T,s)=\{\sigma \in T: s \notin \sigma\}$$

• **Mezclar(T_1,T_2)**: esta operación admite como entrada dos tubos T_1 y T_2 y devuelve como salida un nuevo tubo, $T_1 \cup T_2$, que es la unión de ambos, como multiconjuntos.

• **Detectar(*T*)**: esta es una operación booleana que admite como entrada un tubo T y devuelve como salida **SI**, en el caso en que T contenga algún agregado, y **NO** en caso contrario.

Obsérvese que en el modelo restringido de L. Adleman, la operación molecular **extraer** es la única que implementa el paralelismo masivo en este modelo. Además, en 1995 D. Beaver [5] demostró que el modelo es universal; es decir que tiene la misma potencia computacional que una máquina de Turing.

Seguidamente presentamos un programa molecular que resuelve un problema computacionalmente duro (el problema `3-COL`) usando un tiempo lineal, si bien el espacio usado es exponencial en el tamaño del dato de entrada.

```
Entrada: T (contiene todas las coloraciones del grafo)
        Para i ← 1 hasta p hacer
                T_1 ← +(T, e_i^1(c_1)); T_1^* ← −(T, e_i^1(c_1))
                T_2 ← +(T_1^*, e_i^1(c_2)); T_3 ← −(T_1^*, e_i^1(c_2))
                Para j ← 1 hasta 3 hacer
                        T_j' ← −(T_j, e_i^2(c_j))
                T ← T_1' ∪ T_2'
                T ← T ∪ T_3'
        Detectar(T)
```

4. Programas Moleculares y Sistemas Formales

En esta sección vamos a justificar que los procedimientos mecánicos de los modelos de computación molecular se pueden

considerar como sistemas formales. Para ello se va a introducir una formalización que tiene por objeto caracterizar la verificación formal de programas moleculares diseñados para resolver problemas de decisión, en términos de corrección (o adecuación) y completitud del sistema formal asociado [13].

Teniendo presente que todo problema de optimización se puede describir mediante un problema de decisión (de tal manera que la cantidad de recursos necesarios para esa *traducción* es polinomial), será posible usar una metodología similar para la resolución de problemas de optimización mediante programas moleculares.

Recordemos que un sistema formal consta, básicamente, de:

(a) una *sintaxis*, obtenida a partir de un alfabeto prefijado y de una regla de formación de fórmulas del sistema;
(b) una *semántica* que proporciona el concepto de *validez*; y
(c) una *deducción* que proporciona el concepto de demostrabilidad de una fórmula en el sistema.

Un sistema formal es *adecuado* si y sólo si toda fórmula demostrable es válida, y se dice que es *completo* si y sólo si toda fórmula válida es demostrable.

Informalmente, un problema de decisión es un problema abstracto que sólo admite como respuesta o salida `sí` o `no`. Formalmente, un *problema de decisión*, X, es un par ordenado (E_X, θ_X), en donde E_X es un lenguaje sobre un cierto alfabeto Σ_X (los elementos de E_X se denominan *instancias* o *datos de entrada del problema*) y θ_X es un predicado sobre E_X.

Todo programa molecular que resuelve un determinado problema de decisión se puede representar mediante un sistema formal, en donde:

- La *sintaxis* se define a partir del alfabeto sobre el que se describe el programa, las fórmulas son los tubos de entrada del programa.
- La *semántica* se define a través del problema de decisión.
- La *deducción* se materializa mediante la ejecución del programa.

En este contexto, determinar la validez de una fórmula del sistema formal asociado a un programa molecular que resuelve un problema de decisión, equivale a determinar si dado un tubo de entrada del programa, la respuesta del problema con ese tubo de entrada es `sí`. Asimismo, determinar si una fórmula del sistema es demostrable, equivale a probar que dado un tubo, la salida del programa con ese tubo de entrada es `sí`.

La *verificación formal* de un programa que resuelve un problema abstracto requiere:

- Un *método de especificación formal* que proporcione los criterios para considerar correcto un algoritmo.
- Una *semántica formal del lenguaje de programación* que permita considerar al algoritmo como objeto matemático.
- Unas *reglas formales de demostración* que proporcionen los medios para establecer la veracidad del teorema de corrección (o adecuación) buscado.

En particular, para verificar formalmente un programa molecular diseñado para resolver un problema de decisión hay que probar:

(a) Toda molécula del tubo de salida codifica una solución correcta del problema (*corrección o adecuación del programa molecular*).
(b) Toda molécula del tubo inicial que codifica una solución correcta del problema debe aparecer en el tubo de salida (*completitud del programa molecular*).

Sea P un programa molecular diseñado para resolver un problema de decisión, $X = (E_X, \theta_X)$, y sea Σ_P el alfabeto de entrada de dicho programa. Para cada instancia $a \in E_X$ del problema, notaremos $I(P,a)$ el conjunto de tubos iniciales del programa P relativos al dato de entrada a. Para cada instancia del problema, $a \in E_X$, y cada tubo inicial, $T \in I(P,a)$, notaremos $P(a,T)$ el resultado de ejecutar el programa molecular P con dato de entrada a y tubo de ensayo inicial T.

Definición 2 *Un sistema molecular es un par ordenado (X,P), en donde X es un problema de decisión y P un programa molecular. Verificar un sistema molecular (X,P) consiste en demostrar que el programa molecular P resuelve el problema de decisión X.*

Por tanto, para verificar un sistema molecular (X,P) hay que probar lo siguiente:

- Para cada instancia del problema $a \in E_X$ y cada tubo inicial $T \in I(P,a)$ tales que $P(a,T)=1$, se tiene que $X(a)=1$; es decir, que la salida del programa coincide con la respuesta del problema (*corrección* o *adecuación* del programa P para el problema X).

- Para cada instancia del problema $a \in E_X$ y cada tubo inicial $T \in I(P,a)$ tales que $X(a)=1$ se tiene que $P(a,T)=1$; es decir, que la respuesta del problema coincide con la salida del programa (*completitud* del programa P respecto del problema X).

Definición 3 *Diremos que un sistema molecular (X,P) está verificado si y sólo si para cada instancia $a \in E_X$ y cada tubo $T \in I(P,a)$ se tiene que $P(a,T)=1$ si y sólo si $X(a)=1$.*

Seguidamente vamos a asociar a cada sistema molecular un sistema formal, de tal manera que la verificación de un sistema molecular equivalga a establecer la adecuación y completitud del sistema formal asociado.

Definición 4 *Sea (X,P) un sistema molecular. El sistema formal asociado a (X,P), que notaremos por S(X,P), consta de:*

(a) *Un alfabeto, $\Sigma=\Sigma_X \cup \Sigma_P$.*

(b) *Un conjunto de fórmulas: una fórmula es un par ordenado (a,T), en donde $a \in E_X$ y $T \in I(P,a)$.*

(c) *Un concepto de validez: una fórmula (a,T) es válida si y sólo si $X(a)=1$.*

(d) *Un concepto de deducción: una fórmula (a,T) es demostrable si y sólo si P(a,T)=1.*

Proposición 5 *Sea (X,P) un sistema molecular. Se verifica:*

1. *El sistema formal S(X,P) es adecuado si y sólo si*

$$\forall a \in E_X \forall T \in I(P,a)(P(a,T) = 1 \Rightarrow X(a) = 1)$$

2. *El sistema S(X,P) es completo si y sólo si*

$$\forall a \in E_X \forall T \in I(P,a)(X(a) = 1 \Rightarrow P(a,T) = 1)$$

Corolario 6 *Son equivalentes las proposiciones siguientes:*

- *El sistema molecular (X,P) está verificado.*
- *El sistema formal asociado, S(X,P), es adecuado y completo.*

5. Sistemas Celulares con Membranas

Hasta ahora hemos hablado de modelos de computación natural en los que se simula el modo en que la naturaleza *calcula* a un nivel molecular. A continuación, vamos a pasar a analizar el *nivel celular* desde un punto de vista computacional.

El comportamiento de una célula puede ser considerado como el de una máquina que realiza procesos de cálculo: a través de una distribución jerárquica de membranas interiores se produce el flujo y alteración de las componentes químicas que la propia célula procesa.

La célula puede considerarse como la unidad fundamental de todo organismo vivo, debido su capacidad para realizar una serie de procesos que suelen ser tipificados como característicos de la idea de *vida*:

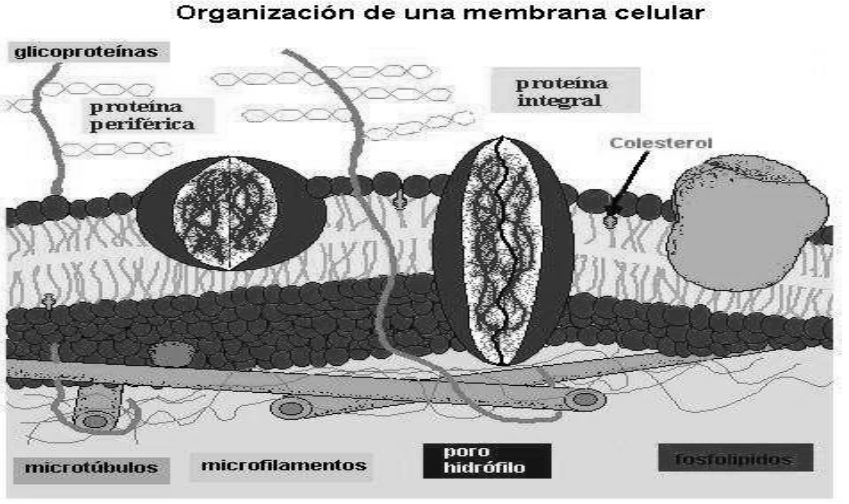
- La replicación del ADN.
- La producción de energía.
- La síntesis de proteínas.
- El desarrollo de procesos metabólicos.

Las células poseen una estructura compleja y, a la vez, muy organizada. En ella se permite ejecución simultánea de múltiples reacciones químicas y en diferentes compartimentos.

Existen dos tipos de células:

- *Procariotas*: que carecen de un núcleo bien definido (y son propias de los organismos unicelulares).
- *Eucariotas*: que poseen un núcleo rodeado por una doble membrana (y son específicas de animales y plantas).

La estructura celular que se considera está inspirada en el *modelo de mosaico fluido* de Singer y Nicholson (1973):

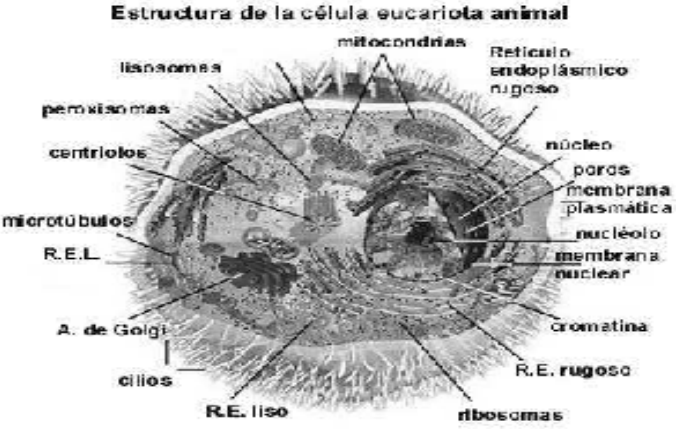


Los procesos que se producen en una célula son lo suficientemente complejos como para que sea imposible modelizarlos completamente ya que un modelo de computación que intente simular *literalmente* esos procesos dejará de ser práctico, salvo desde un punto de vista biológico. Con este tipo de computación, se pretende crear un modelo abstracto que simule, de la forma más aproximada posible, el comportamiento de las células y que nos permita (al menos en principio) obtener soluciones alternativas de problemas computacionalmente intratables desde un punto de vista no convencional. Para ello, hay que hacer emerger todas aquellas

características del comportamiento y constitución de la célula que puedan ser de utilidad para la elaboración de un modelo de computación que sea a la vez potente (en cuanto a los problemas que pueda resolver) y sencillo (en cuanto a su definición, implementación y ejecución).

Las partes más importantes de una célula eucariota son las siguientes:

- Una especie de piel, denominada *membrana plásmica*.
- El corazón de la célula (denominado *núcleo*) cuyo objetivo fundamental es almacenar el ADN (identificador de la célula).
- El resto de la célula, denominado *citoplasma*, que contiene:
  - La *mitocondria*, que se encarga de producir energía.
  - El *aparato de Golgi* que es una especie de fábrica de proteínas y juega un papel esencial en el metabolismo de la célula
  - Los *ribosomas* que se encargan de la síntesis de las proteínas.
  - El *retículo endoplásmico* que es una red de membranas interconectadas y permite el transporte de moléculas y comunicación entre los distintos compartimentos.
  - Los *lisosomas* que se consideran como los *estómagos* de las células



Las membranas biológicas están involucradas en la mayoría de las reacciones químicas que tienen lugar en las células.

- Son canales selectivos de comunicación (barreras *semipermeables*) entre los distintos compartimentos, así como entre la propia célula y su entorno.
- Controlan un flujo de datos; es decir, de *información* tanto en el interior de la célula como entre ésta y su entorno.

De la importancia y actualidad de las membranas biológicas baste decir que el Premio Nobel de Química de 2003 se le concedió conjuntamente a P. Agre y R. MacKinnon por sus investigaciones sobre los canales proteínicos de las membranas.

En particular, la membrana *plásmica* separa el espacio interno de la célula protegiéndolo de su entorno. Las restantes membranas (denominadas *internas*) proporcionan la estructura jerarquizada propia de la célula y ofrecen la adecuada protección al núcleo, que es el depositario de la información genética. Las membranas están involucradas de manera decisiva en la mayoría de reacciones químicas que tienen lugar en los compartimentos celulares y pueden considerarse como barreras semipermeables que, o bien permiten el paso de algunas sustancias químicas en cualquiera de los dos sentidos (dentro–fuera o fuera–dentro), o bien bloquean el paso de sustancias. Las membranas se comportan como canales selectivos de comunicación para la transferencia de compuestos químicos entre distintos compartimentos, así como entre la propia célula y su entorno, controlando así un cierto flujo de datos (es decir, de información).

Es fácil encontrar unas primeras similaridades entre células y máquinas

- En una célula viva:
  - Cada *membrana* trabaja con *compuestos* químicos de acuerdo con unas *reacciones* específicas.
- En una máquina paralela:
  - Cada *procesador* trabaja con *datos* de acuerdo con un *programa* específico.

Teniendo presente que muchos desarrollos en Informática, se han producido bajo la inspiración de procesos que se dan en la naturaleza, cabe realizarse algunas preguntas, al respecto:

- ¿Es la *célula* una fuente de inspiración para la informática?
- ¿Es posible definir un modelo de computación inspirado en las *células*?
- ¿Es posible implementar computaciones a través de las *células* vivas?

6. Computación Celular con Membranas

La Computación Celular con Membranas es una disciplina emergente de la Computación Natural que fue considerada en octubre de 2003 por el Thomson Institute for Scientific Information (**ISI**) como una *Fast Emerging Research Front in Computer Science* [41]. Fue introducida por Gheorghe Păun a finales de 1998 y el artículo fundacional [32] fue nominado por el ISI como el *Fast Breaking Paper* [42] de febrero de 2003. Desde entonces ha recibido una atención importante de la comunidad científica internacional. El interés suscitado por esta disciplina queda patente al constatar que, en poco más de nueve años, se han publicado cerca de 1200 artículos, con la participación de más de 300 autores, 3 libros, 22 volúmenes colectivos y 24 tesis doctorales. Para más detalle ver la página [43], mantenida en Viena bajo los auspicios del European Molecular Computing Consortium.

En las ideas originales de Gh. Păun, los sistemas celulares con membranas no fueron introducidos propiamente para modelizar completamente la estructura y funcionamiento de una célula, sino más bien para analizar algunos hechos computacionalmente relevantes que pueden ser abstraídos de las mismas.

En esta sección vamos a dar una descripción informal de los sistemas celulares básicos de transición.

Una *estructura de membranas* consta de un conjunto finito de membranas que se encuentran dentro de una membrana principal (denominada *piel*) y delimita una serie de *regiones* o compartimentos formados por el espacio comprendido entre una membrana y las

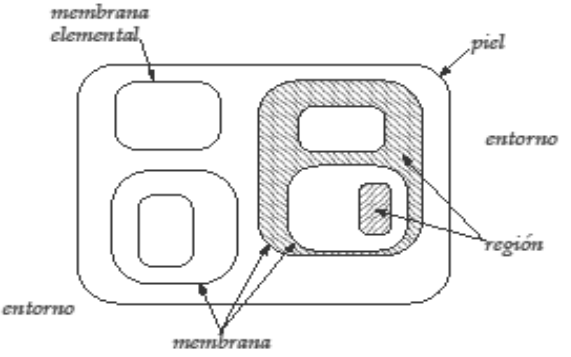
situadas inmediatamente dentro de ella (si las hay). Una membrana es *elemental* si no contiene ninguna membrana.

Una estructura de membranas puede ser descrita como una cadena sobre un alfabeto finito y se puede representar mediante diagramas de Venn a través una familia de subconjuntos de un conjunto base, tal que dos conjuntos cualesquiera de la familia o son disjuntos, o uno de ellos está contenido en el otro. También se puede considerar como un árbol cuya raíz es la membrana piel y las hojas son las membranas elementales.

Las sustancias químicas contenidas en las distintas regiones de las membranas celulares se representan mediante símbolos de un alfabeto finito. Por tanto, los objetos son símbolos atómicos, indivisibles, sin *estructura interna*, pero que pueden *evolucionar* a lo largo del tiempo.

Teniendo presente que las diferentes sustancias químicas pueden aparecer en una región en diferentes cantidades, representaremos el contenido de cada región a través de *multiconjuntos* de objetos (conjuntos cuyos elementos pueden aparecer repetidos, es decir, con multiplicidades asociadas a cada uno de ellos).

Las reacciones químicas que se producen en las regiones se representan a través de *reglas de evolución* que, básicamente, son reglas de reescritura de un sistema formal que actúan sobre los objetos modificándolos y/o trasladándolos de un compartimento a otro de acuerdo con una determinada semántica.



Una *configuración* de un sistema celular consta de una estructura de membranas y una familia de multiconjuntos de objetos asociados a cada región. En la descripción de un tal sistema, aparece

siempre una configuración denominada *inicial*. En algunas variantes existe una *membrana de entrada* que permite añadir un multiconjunto de objetos antes de que se produzcan evoluciones.

Las reglas de evolución serán aplicadas de manera *no determinista, paralela y maximal*; es decir, las reglas a ser usadas y los objetos involucrados en la misma serán escogidos de tal manera que en cada paso de computación todos los objetos que *puedan* evolucionar *tienen* que evolucionar. Los objetos pueden pasar de una región a otra a través de las membranas y éstas pueden ser disueltas, creadas o divididas. Así se produce una *transición* de una configuración a otra.

Una *computación* en un sistema celular es una sucesión (finita o no) de configuraciones tal que la primera de ellas es una configuración inicial y toda configuración de la sucesión que no sea la inicial se obtiene de la anterior mediante un paso de transición. Se dice que una configuración es de *parada* si no existe ninguna regla de evolución que sea aplicable a dicha configuración. Se dice que una *computación* es de *parada* si contiene una configuración de parada, y suele tener asociado un *resultado* codificado en una membrana elemental distinguida (denominada de *salida*) de la correspondiente configuración de parada.

Así pues, una computación en un sistema celular parte de una configuración y mediante sucesivas transiciones, o bien se llega a una situación de parada (en tanto que el sistema no puede seguir evolucionando), o bien, el sistema nunca se detiene (en tanto que sigue evolucionando de manera indefinida).

Admitiremos que toda computación ejecuta un proceso sincronizado; es decir, se supone que hay una especie de *reloj universal* que marca las actuaciones de todos los elementos que integran el sistema celular. Hay que destacar la existencia de dos niveles de paralelismo en la ejecución: por una parte, las reglas asociadas a una membrana son aplicadas de manera simultánea y, por otra, estas operaciones son realizadas en el mismo instante en todas las membranas que vertebran el sistema. Es decir, en un paso de reloj se ejecutarán todas las reglas que pueden ser aplicadas (de manera maximal), y en todas las membranas.

Los sistemas celulares con membranas constituyen un modelo de computación que conjuga la elegancia y sencillez del modelo, con su proximidad a la realidad de la célula, desde un punto de vista

biológico, abstrayendo de la estructura y funcionamiento de las células una serie de ideas computacionalmente relevantes. Son modelos matemáticos con propiedades atractivas desde el punto de vista computacional: algunos sistemas celulares son *completos* (tienen la misma potencia computacional que las máquinas de Turing) y *eficientes* (capaces de proporcionar soluciones polinomiales de problemas **NP**–completos).

Aunque los sistemas celulares tienen una inspiración biológica, se debe resaltar el hecho de que constituyen igualmente un buen modelo teórico de computación distribuida, en donde distintas unidades de cálculo trabajan independientemente pero estructuradas en una cierta jerarquía vertical. Por ejemplo, la jerarquización usada en las conexiones establecidas en redes de ordenadores como Internet puede ser representada como una estructura de membranas, en la que los nodos de la red se interpretan como las membranas que forman el sistema, y el flujo de información entre nodos conectados se interpreta como las componentes químicas que dichas membranas generan e intercambian.

Asimismo, sistemas dinámicos complejos, como aquellos que surgen en el estudio de la dinámica de poblaciones, pueden ser también interpretados como sistemas celulares en los que los individuos de las distintas especies interaccionan entre sí dentro de hábitats que permiten el traspaso de los mismos.

Como primera nota diferenciadora con respecto a otros modelos de computación natural (por ejemplo, los modelos de computación molecular basados en ADN) conviene hacer notar que no está descrito a través de un lenguaje de programación; es decir, no proporciona una serie de operaciones básicas susceptibles de ser secuenciadas sobre un dato de entrada para obtener un resultado final (solución del problema que se pretende resolver), sino que se generan *dispositivos* (al modo de las máquinas de Turing) cuya ejecución modifica el contenido de las distintas componentes que los integran hasta llegar, en su caso, a un estado de parada (en el que la máquina deja de funcionar).

En resumen, los ingredientes sintácticos básicos de un sistema P son los siguientes:

• Un *alfabeto finito*, cuyos elementos se denominan *objetos*.

- Una *estructura de membranas* que delimitan *regiones*.
- Una serie de multiconjuntos de objetos asociados a las distintas regiones que determinan la estructura de membranas.
- Un conjunto de *reglas de evolución* y de prioridades asociados a cada región.
- Dos membranas distinguidas: una de *entrada* y otra de *salida*.

Los ingredientes semánticos principales de un sistema P son los conceptos de:

- *Configuración*.
- *Transición* de una configuración a otra.
- *Computación* a partir de una configuración inicial.

Para verificar formalmente una familia *F* de sistemas P diseñada para resolver un problema de decisión hay que probar dos propiedades:

(a) Si **una** computación del sistema P que procesa una determinada instancia del problema es de aceptación (es decir, devuelve `sí`), entonces la respuesta del problema a esa instancia es `sí` (*corrección o adecuación del sistema P*).
(b) Si la respuesta del problema a una cierta instancia es `sí`, entonces **toda** computación del sistema P que procesa esa instancia es de aceptación (es decir, devuelve `sí`) (*completitud del sistema P*).

En tal situación, diremos que la familia *F resuelve* el problema *X*.

Obsérvese que todo sistema P de una familia que resuelve un problema de decisión es **confluente**, en el sentido de que cualquier computación del mismo, asociada a un dato de entrada, proporciona la misma respuesta. En consecuencia, dichos dispositivos computacionales son *fiables* a pesar de ser no–deterministas; es decir, capturan el *verdadero concepto de algoritmo*.

Definición 7 *Un sistema celular es un par ordenado (X,F), en donde X es un problema de decisión y F es una familia de sistemas P de tal manera que cada instancia del problema determina unívocamente un sistema de la familia. Verificar un sistema celular*

(X,F) consiste en demostrar que la familia F resuelve el problema de decisión X.

Por tanto, para verificar un sistema celular (*X*,*F*) hay que probar lo siguiente:

- Para cada instancia del problema $a \in E_X$, si existe una computación de aceptación del sistema que procesa esa instancia, entonces $X(a)=1$ (*corrección* o *adecuación* de la familia *F* respecto del problema *X*).
- Para cada instancia del problema $a \in E_X$, si $X(a)=1$ entonces toda computación del sistema que procesa esa instancia es de aceptación (*completitud* de la familia *F* para el problema *X*).

Definición 8 *Diremos que un sistema celular (X,F) está verificado si y sólo si para cada instancia* $a \in E_X$ *se tiene que X(a)=1 si y sólo si toda computación del sistema que procesa la instancia a es de aceptación.*

Seguidamente vamos a asociar a cada sistema celular un sistema formal, de tal manera que la verificación de un sistema celular equivalga a establecer la adecuación y completitud del sistema formal asociado.

Definición 9 *Sea (X,F) un sistema celular. El sistema formal asociado a (X,F), que notaremos por S(X,F), consta de:*

(a) *Un alfabeto,* $\Sigma=\Sigma_X \cup \Sigma_F$*, en donde* Σ_F *es la unión de los alfabetos de trabajo de los sistemas de la familia F.*

(b) *Un conjunto de fórmulas: una fórmula es un par ordenado (a,C), en donde* $a \in E_X$ *y C es una computación del sistema que procesa a.*

(c) *Un concepto de validez: una fórmula (a,C) es válida si y sólo si X(a)=1.*

(d) *Un concepto de deducción: una fórmula (a,C) es demostrable si y sólo si la computación C es de aceptación.*

A continuación se establece que verificar un cierto sistema celular equivale a probar que el sistema formal asociado es adecuado y completo.

Proposición 10 *Sea (X,F) un sistema celular. Se verifica:*

1. El sistema formal S(X,F) es adecuado si y sólo si para cada instancia $a \in E_X$ *y cada computación C del sistema que procesa a se tiene que si C es de aceptación, entonces X(a)=1.*

2. El sistema S(X,F) es completo si y sólo si para cada instancia $a \in E_X$ *tal que X(a)=1 se tiene que toda computación C del sistema de F que procesa la instancia a es de aceptación.*

Corolario 11 *Son equivalentes las proposiciones siguientes:*

- *El sistema celular (X,F) está verificado.*
- *El sistema formal asociado, S(X,F), es adecuado y completo.*

7. Algunas Aplicaciones y Retos de la Computación Celular

Vamos a finalizar citando algunas de las aplicaciones más interesantes que se han desarrollado hasta la fecha en el área de la Computación Celular con Membranas y proponiendo algunos retos importantes en dicha disciplina.

Entre las aplicaciones desarrolladas destacamos las siguientes:

- Simulación eficiente de máquinas de Turing deterministas mediante sistemas P básicos de transición [36].
- Resolución eficiente de problemas *computacionalmente duros* [19],[20].
- Caracterizaciones de la conjetura **P≠NP** en un marco de computación no convencional [37], [29].

- Modelado y simulación computacional de procesos biológicos:
  - Rutas señalizadoras que intervienen en la génesis de tumores (interacción entre las proteínas p53 y MdM2, estudio del factor de crecimiento epidérmico EGFR, apoptosis mediatizada por la proteína FAS, etc.) [38], [11], [26], [31].
  - Comunicación inteligente de bacterias (Quorum Sensing en la bacteria Vibrio Fischeri) [39].
  - Sistemas de regulación de genes (el Lac Operon) [40].
  - Estudio de la dinámica global de ciertos tumores cancerígenos [22].
- Desarrollo de herramientas hardware para la simulación de sistemas P en ordenadores electrónicos [12], [18], [6].
- Simulación de arquitecturas paralelas [8], [9].
- Diseño de dispositivos computacionales inspirados en los sistemas P y que trabajan a modo de tejidos [14], [21].
- Especificaciones y modelado de ciertos modelos económicos [33].
- Computación neuronal de tercera generación basada en descargas eléctricas (*spikes*) que permite el tratamiento de sucesiones arbitrariamente grandes de bits [10], [34], [35].
- Modelado y simulación computacional de la dinámica de poblaciones (se ha realizado un estudio de la evolución de una población de quebrantahuesos en la zona pirenaica catalana) [7].
- Simulación del comportamiento de sociedades (colonias de abejas) [17].

Entre los retos más importantes que tiene hoy día la Computación Celular con Membranas destacamos los siguientes:

- Implementación de dispositivos computacionales de alguna variante de sistemas P.
- Elaboración de un lenguaje de programación celular que sirva de asistente en la tarea de diseño de modelos.
- Diseño de estrategias de búsqueda eficiente de *buenas* computaciones en sistemas celulares
- Desarrollo de sistemas distribuidos para modelos celulares *cell–like*, *tissue–like* y *spiking neural P systems*.

• Establecimiento de nuevas fronteras entre la tratabilidad e intratabilidad de problemas.
• Modelado y simulación de sistemas dinámicos complejos.
• Formalización de los sistemas de producción.
• Representación y tratamiento de fractales.
• Simulación de sistemas de razonamiento automático y Desarrollo de técnicas de aprendizaje automático.
• Formulación de hipótesis (predicciones) acerca de procesos biológicos en el marco de la Biología de Sistemas.

Agradecimientos

Este trabajo ha sido subvencionado por el proyecto de Excelencia de la Junta de Andalucía TIC–581 y el proyecto I+D+I del Ministerio de Educación y Ciencia TIN2006–13425.

Referencias

1. L. ADLEMAN. Molecular Computation of Solutions to Combinatorial Problems, *Science*, 268, November 1994, 1021–1024.

2. L. ADLEMAN. On constructing a molecular computer. En R.J. Lipton y E.B. Baum (eds.) *DNA based computers*, American Mathematical Society, 1996, 1–22.

3. M. AMOS, S. WILSON, D.A. HODGSON, G. OWENSON, A. GIBBONS. Practical implementation of DNA computations, in *Unconventional Models of Computation*, Springer, 1998, 1–18.

4. M. AMOS. *DNA computing*. Ph. D. Thesis, University of Warwick, 1997.

5. D. BEAVER. A universal molecular computer. En *DNA based computers* (R.J. Lipton and E.B. Baum, eds.), American Mathematical Society, 1996, 29–36.

6. R. BORREGO, D. DÍAZ, M.J. PÉREZ-JIMÉNEZ. Tissue simulator: A graphical tool for tissue P systems. En Gy. Vaszil (ed.) *Proceedings of the International Workshop Automata for Cellular and Molecular Computing, MTA SZTAKI*, Budapest, Hungary, August 31, 2007, pp. 23–34.

7. M. CARDONA, M.A. COLOMER, M.J. PÉREZ-JIMÉNEZ, D. SANNY, A. MARGALIDA. Modelling Ecosystems using P Systems: The Bearded Vulture, a case study.

Aceptado en el Nineth Workshop on Membrane Computing, WMC-9, Edinburgh (UK), July 28-31, 2008.

8. R. CETERCHI, M.J. PEREZ-JIMENEZ. On simulating a class of parallel architectures. *International Journal of Foundations of Computer Science*, **17**, 1 (2006), 91–110.

9. R. CETERCHI, M.J. PEREZ-JIMENEZ. A.I. TOMESCU. Simulating the bitonic sort using P systems. *Lecture Notes in Computer Science*, **4860** (2007), 172-192.

10. H. CHEN, R. FREUND, M. IONESCU, GH. PAUN, M.J. PEREZ-JIMENEZ. On string languages generated by Spiking Neural P systems. *Fundamenta Informaticae*, **75**, 1-4 (2007), 141–162.

11. S. CHERUKU, A. PĂUN, F.J. ROMERO, M.J. PÉREZ-JIMÉNEZ, O.H. IBARRA. Simulating FAS-induced apoptosis by using P systems. *Progress in Natural Science*, **17**, 4 (2007), 424–431.

12. A. CORDÓN, M.A. GUTIERREZ, M.J. PÉREZ-JIMÉNEZ, A. RISCOS N. Cellular solutions to some numerical NP-complete problems: A Prolog implementation. En M. Gheorghe (ed.) *Molecular Computational Models: Unconventional Approaches*, Idea Group Publishing, 2005, chapter V, pp. 115–149.

13. D. DÍAZ, M.J. PÉREZ-JIMÉNEZ, A. J. RAMOS. Molecular programs as formal systems. En E. Alvarez, R. Bosch, L. Villamil (eds.) *12th International Congress of Logic, Methodology and Philosophy of Science. Volume of abstracts*, Oviedo, Spain, August 7–13, 2003, pp. 160–161.

14. D. DÍAZ, M.A. GUTIÉRREZ, M.J. PÉREZ-JIMÉNEZ, A. RISCOS N. A linear-time tissue P system based solution for the 3-coloring problem. *Electronic Notes in Theoretical Computer Science*, **171** (2007), 81-93.

15. R.P. FEYNMAN. There's plenty of room at the bottom. En D.H. Hilbert (ed.) *Miniaturization*, Reinhold, 1961, 282–296.

16. R.M. GAREY, D.S. JOHNSON. *Computers and intractability*, W.H. Freeman and Company, New York, 1979.

17. M. GHEORGHE, C. MARTÍN-VIDE, V. MITRANA, M.J. PÉREZ-JIMÉNEZ. An agent based approach of collective foraging. *Lecture Notes in Computer Science*, **2686** (2003), 638–645.

18. M.A. GUTIÉRREZ, M.J. PÉREZ-JIMÉNEZ, A. RISCOS N. Available membrane computing software. En G. Ciobanu, Gh. Păun, M.J. P? rez-Jim? nez (eds.) *Applications of Membrane Computing*, Natural Computing Series, Springer-Verlag, 2006, Chapter **15**, pp. 411-436.

19. M.A. GUTIÉRREZ, M.J. PÉREZ-JIMÉNEZ, A. RISCOS, F.J. ROMERO. Computational efficiency of dissolution rules in membrane systems. *International Journal of Computer Mathematics*, **83**, 7 (2006), 593–611.

20. M.A. GUTIÉRREZ, M.J. PÉREZ-JIMÉNEZ, A. RISCOS, F.J. ROMERO, A. ROMERO. Characterizing tractability by cell-like membrane systems. En K.G. Subramanian, K. Rangarajan, M. Mukund (eds.) *Formal models, languages and applications*, World Scientific, Series in Machine Perception and Artificial Intelligence - Vol. 66, 2006, chapter **9**, pp. 137–154.

21. M.A. GUTIÉRREZ, M.J. PÉREZ-JIMÉNEZ, A. RISCOS, F.J. ROMERO. Cell-like and tissue-like membrane systems as recognizer devices. En D.A. Pelta, N. Krasnogor (eds.) *Proceedings of the Workshop on Nature Inspired Cooperative Strategies for Optimization, NICSO 2006*, June 29-30, 2006, Granada, Spain, 173-186.

22. M.A. GUTIÉRREZ, M.J. PÉREZ-JIMÉNEZ, A. RISCOS, F.J. ROMERO. How to express tumours using membrane systems. *Progress in Natural Science*, **17**, 4 (2007), 449–457.

23. J. HARTMANIS, P.M. LEWIS, R.E. STEARN. Hierarchies of memory limited computations. *Proceedings of the Sixth Annual IEEE Symposium on Switching Circuit Theory and Logical Design*, 1965, 179–190.

24. T.H. HEAD. Formal language theory and DNA: an analysis of the generative capacity of specific recombinant behaviors. *Bulletin of Mathematical Biology*, **49**, 1987, 737–759.

25. J.H. HOLLAND *Adaptation in natural and artificial systems*. MIT Press, 1975.

26. J. JACK, F.J. ROMERO, M.J. PÉREZ-JIMÉNEZ, O.H. IBARRA, A. PĂUN. Simulating apoptosis using discrete methods: A membrane system and a stochastic approach. En M. Domaratzki and K. Salomaa (eds.) *Proceedings of the Language Theory in Biocomputing Workshop, satellite of the Unconventional Compution, UC'07*, Kingston, Ontario, Canada, August 15, 2007, pp. 50–63.

27. P.M. LEWIS, R.E. STEARN, J. HARTMANIS. Memory bounds for recognition of context-free and context-sensitive languages. *Proceedings of the Sixth Annual IEEE Symposium on Switching Circuit Theory and Logical Design*, 1965, 191–202.

28. R.J. LIPTON. DNA Solution of Hard Computational Problems, *Science*, 268, April 1995, 542–545.

29. G. MAURI, M.J. PEREZ-JIMENEZ, C. ZANDRON. On a Paun's conjecture in membrane systems. *Lecture Notes in Computer Science*, **4527** (2007), 180–192.

30. W.S. MCCULLOCH, W. PITTS. A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics*, vol. 5, 1943, 115–133.
31. A. PĂUN, M.J. PÉREZ-JIMÉNEZ, F.J. ROMERO. Modeling signal transduction using P systems. *Lecture Notes in Computer Science*,**4361** (2006), 100–122.

32. GH. PĂUN. Computing with membranes. *Journal of Computer and System Sciences*, 61, 1 (2000), 108–143, y *Turku Center for Computer Science-TUCS Report* No. 208 (1998) (www.tucs.fi).

33. GH. PĂUN, R. PĂUN. Membrane computing as a framework for modeling economic processes, *Proceedings of the Seventh International Symposium on Symbolic and Numeric Algorithms for Scientific Computing (SYNASC 2005)*, IEEE Computer Society, 2005, 11–18.

34. GH. PĂUN, M.J. PÉREZ-JIMÉNEZ, GR. ROZENBERG. Computing morphisms by spiking neural P systems. *International Journal of Foundations of Computer Science*, **18**, 6 (2007), 1371–1382.

35. GH. PĂUN, M.J. PÉREZ-JIMÉNEZ, A. SALOMAA. Spiking neural P systems: An early survey. International Journal of Foundations of Computer Science, **18**, 3 (2007), 435–455.

36. M.J. PÉREZ-JIMÉNEZ, A. ROMERO JIMÉNEZ. Simulating Turing Machines by P systems with External Output. *Fundamenta Informaticae*, **49**, 1–3 (2002), 273–287.

37. M.J. PÉREZ-JIMÉNEZ, A. ROMERO, F. SANCHO. The P versus NP problem through cellular computing with membranes. En N. Jonoska, Gh. Paun, Gr. Rozenberg (eds.) Aspects of Molecular Computing, Essays dedicated to Tom Head on the ocassion of his 70th birthday, *Lecture Notes in Computer Science*, **2950** (2004), 338–352.

38. M.J. PÉREZ-JIMÉNEZ, F.J. ROMERO. A study of the robustness of the EGFR signalling cascade using continuous membrane systems. *Lecture Notes in Computer Science*, **3561** (2005), 268–278.

39. F.J. ROMERO, M.J. PÉREZ-JIMÉNEZ. A model of the Quorum Sensing System in Vibrio Fischeri using P systems. *Artificial Life*, 14, 1 (2008), 95-109.

40. F.J. ROMERO, M.J. PÉREZ-JIMÉNEZ. Modelling gene expression control using P systems: The Lac Operon, a case study. *BioSystems*, **91**, 3 (2008), 438–457.

41. ISI web page October 2003: http://esi-topics.com/erf/october2003.html

42. ISI web page February 2003: (http://esi-topics.com/fbp/fbp-february2003.html)

43. P systems web page http://ppage.psystems.eu/

PARADIGMA, ACTO E SENTIDO EM *Sobre a lógica e a teoria da ciência* DE JEAN CAVAILLÈS

Nuno Miguel Proença

Gostaria de começar por agradecer o convite que me foi feito pela Profª Olga Pombo e do Prof. Nepomuceno para participar neste colóquio sobre *Lógica e filosofia da Ciência*. Apesar de não ser a área na qual trabalhei durante mais tempo até agora, não deixa de ser um daqueles que me interessa desde há anos. Peço-vos por isso que desculpem alguma imprecisão da minha parte. A minha presença aqui hoje convosco deve-se ao facto de há algum tempo ter podido estudar e depois traduzir para português o último escrito de Jean Cavaillès intitulado *Sobre a lógica e a teoria da Ciência.*

Permitam-me então que diga algumas palavras sobre a obra antes de vos apresentar, ou de reapresentar, para quem já as conhece, as noções de *Paradigma e de temático,* que são aquelas que, junto com as noção de *acto e sentido* dão conta, em parte, da unidade e da autonomia da geração conceptual que caracteriza o desenvolvimento das ciências demonstrativas, segundo Cavaillès, e a estrutura lógica das mesmas.

O livro é denso e de leitura exigente, por requerer, como escreve Bachelard no prefácio à segunda edição, um esforço da parte dos leitores para se familiarizarem com os desenvolvimentos da lógica e da matemática dos finais do século XIX e inícios do século XX e também com a filosofia das ciências deste período. Acresce talvez à dificuldade o facto do livro ser considerado incompleto por lhe faltar uma introdução que Cavaillès nunca redigiu. O título foi-lhe dado por Charles Ehresman e Georges. Canguilhem, os primeiros editores, depois do desaparecimento do autor no decurso das suas actividades de resistente anti-nazi durante a segunda guerra mundial. Em França, Cavaillès foi «um dos quatro ou cinco elementos fundadores dos primeiros movimentos de Resistência. Neste domínio, a tarefa de Cavaillès foi esmagadora até à segunda detenção que teve lugar em Agosto de 1943 e foi definitiva, escreve Canguilhem. «Esta tarefa, no entanto, não lhe afastou o olhar do trabalho filosófico. Na verdade

para ele ambos estavam ligados. Voltar à reflexão parecia-lhe indispensável, no seio da acção, para garantir o sentido desta. Tinha a intenção, da qual nos tinha dado conta várias vezes, de escrever um *Tratado de Lógica*, mas costumava acrescentar rindo: «Não terei tempo de o escrever senão na prisão.» Foi o que de facto aconteceu. Ao embarcar para Inglaterra, numa noite de Agosto de 1942, Cavaillès foi detido numa praia das vizinhanças de Narbonne e foi transferido para a prisão militar de Montpellier. Foi então que, para vencer o desespero do cativeiro, empreendeu a redacção do trabalho cujo projecto tinha nutrido»[1]. No decurso do ano de 1942 o tribunal militar de Montpellier encerrou o caso Cavaillès por considerar que se tratava de uma improcedência jurídica. Mas um mandato de captura foi emitido contra o filósofo que por essa razão foi enviado «para o campo de Saint-Paul d' Eyjaux, perto de Limoges. Aí permaneceu, se é que podemos dizer, só o tempo que foi preciso para terminar o trabalho e evadiu-se, com o manuscrito no bolso, nos primeiros dias do ano de 1943. Já só lhe faltava, segundo dizia, fazer referência aos autores em rodapé, antes de dar o manuscrito ao editor. Foi este derradeiro trabalho de acerto que as circunstâncias o impediram de levar a cabo. Em Março de 1943, Cavaillès consegue partir para a Inglaterra. Volta de lá em Maio. Em Junho, escapa por pouco à detenção pela polícia alemã. Em Agosto é detido pelos serviços de contra-espionagem alemães» e fuzilado «entretanto, tinha entregue por precaução à irmã, a senhora Ferrières, um exemplar dactilografado do trabalho, por ele corrigido»[2], que os editores publicaram em 1946 e cuja tradução para Português propomos agora.

O título foi dado à obra pelos mesmos e, apesar de não querer impor um programa de leitura, acaba por resumir a preocupação do autor em saber se se pode formular uma teoria da ciência que tenha por modelo a lógica e a matemática. O percurso argumentativo de Cavaillès é delimitado em grande parte pela reflexão sobre o trabalho de quatro filósofos: Kant, Bolzano, Carnap e Husserl. A série das discussões levadas a cabo conclui com a frase que deu que pensar a muitos dos leitores de Cavaillès: «Não é uma filosofia da consciência mas uma filosofia do conceito que pode dar uma doutrina da ciência. A necessidade geradora não é a de uma actividade, mas a de uma

[1],G. Canguilhem, Ch Ehresman, *Prólogo dos editores à primeira edição*.
[2] *Idem*.

dialéctica»[3]. No entanto, é uma filosofia da consciência, a de Kant, que inaugura a reflexão, e outra, a de Husserl, que alimenta a terceira e última parte do livro. Entretanto, são duas filosofias da ciência que se fundam na lógica, a de Bolzano e a de Carnap, que marcam os momentos centrais do percurso. «A de Bolzano, escreve Jan Sebestik no posfácio de 1997 a este trabalho de Cavaillès, afirma um domínio de entidades lógicas e matemáticas abstractas, enquanto que a segunda constrói o mundo a partir de uma relação entre dados sensíveis e considera a lógica (e a matemática) como um instrumento sem conteúdo próprio, destinado a descrever o mundo, quer dizer a coordenar os domínios sensíveis para prever outras constelações de dados sensíveis»[4]. Entre ambas as leituras, Cavaillès leva a cabo uma análise das doutrinas de Brouwer e de Brunschvicg no seguimento da análise da posição kantiana.

Cada uma das doutrinas analisadas (Kant, Brouwer, mas também Brunschvig, Bolzano, Carnap, Tarski e Husserl) contribui de maneira essencial, segundo Cavaillès, para a constituição de uma teoria da ciência. Kant, por insistir na actividade construtiva da consciência e na noção de forma; Bolzano por estabelecer uma teoria da ciência que assegura tanto a autonomia do progresso científico – ao substituir a evidência intuitiva pela exigência demonstrativa – como pela construção do seu sistema lógico; Brouwer por conceber a matemática como devir autónomo, «mais acto que doutrina»; Husserl, pela solidariedade efectiva entre objectos e actos, entre a teoria dos juízos (apofântica) e a ontologia formal (*mathesis universalis*), pela sua solução da relação entre matemática e física, como sendo percorridas por um só movimento do pensamento. Por fim pela historicidade do seu olhar sobre a ciência.

As noções de *paradigma*, associada à noção de *tematização* acompanhada por uma reflexão sobre o *acto* e o *sentido* das operações demonstrativas características da lógica e da matemática, aparecem como elementos essenciais de uma caracterização do movimento pelo qual novos seres e novas teorias se constituem no devir conceptual autónomo característico da Ciência. Cavaillès responde assim à hipótese de Bolzano e à posição intuicionista sobre a teoria da ciência.

[3] Cavaillès, *Sobre a Lógica e a Teoria da Ciência*, último parágrafo.
[4] J. Sebestik, *Postface*, in *Sur la Logique et la Théorie de la Science*, Paris, Vrin, 1997, p.93.

Apararem no decurso da obra, retomando posições expressas nos trabalhos anteriores do filósofo, entre uma reflexão sobre as dificuldades da hipótese de uma sintaxe absoluta da ciência e a reflexão sobre o interesse e os limites da posição de Husserl. Gostaria de voltar a percorrer brevemente o percurso argumentativo e problemático do texto que leva à análise destas duas noções que, no meu entender, são elementos determinantes para compreender a originalidade do pensamento da ciência do filósofo matemático, nutrido de poesia que, no dizer de G.Canguilhem[5], citava Rimbaud nas suas lições sobre a experiência e nos ensinou que, antes de ser irmã do sonho, acção tem de ser filha do rigor.

No entender de Cavaillès, «depois da análise kantiana ficam abertas duas possibilidades para a doutrina da ciência: ou se acentua a noção de sistema demonstrativo ou se acentua a noção de *organon* matemático. À primeira está ligada a concepção lógica inaugurada por Bolzano e prosseguida simultaneamente e de diversas maneiras pelos formalistas e por Husserl. À segunda [estão ligadas] as filosofias epistemológicas da imanência, como podem ser designadas a de Léon Brunschwicg ou a de Brouwer»[6]. Estas duas últimas têm em comum o cuidado de opor a imanência do pensamento autêntico à abstracção da linguagem. A apresentação rápida que Cavaillès faz do intuicionismo esclarece de alguma forma a sua própria posição no que diz respeito à relação da lógica com uma teoria da ciência. Passo a citar Cavaillès acerca de Brouwer. O pensamento verdadeiro de Brouwer, «exprimido por exemplo em *Matematische, Wissenschaft und Sprache*»[7] poderia caracterizar-se pelo seguinte: «Em primeiro lugar a matemática é um devir autónomo, «mais um acto do que uma doutrina», cuja definição à partida é impossível mas cujos momentos, na solidariedade necessária que os caracteriza, dão prova de uma essência original. Da díade às teorias elaboradas, há continuidade e imprevisibilidade»[8]. Mas este acto gerador e critério do que é matematicamente verdadeiro deve situar-se concretamente, submeter-se a condições de cumprimento efectivo. É interessante o parentesco entre esta posição e a que Cavaillès exprime numa conferência publicada depois da sua

[5] G. Canguilhem, *In Memoriam*.
[6] Cavaillès, *Sobre a Lógica…* §13.
[7] *Ibid.*
[8] *Ibid.* §14.

morte e intitulada *O pensamento matemático*. Nesta, em conjunto com Lautman, Cavaillès caracteriza a actividade «dos matemáticos como uma actividade experimental em que a experiência é um sistema de gestos, regidos por uma regra e submetido a condições independentes desses gestos»[9]. «Se desta forma não existe método no sentido tradicional, existe dominação do decurso da ciência por regras»[10]. Ora, «se as regras são modificáveis a cada instante, segundo a teoria à qual se aplicam, têm um princípio comum, a exigência de uma constructibilidade positiva, a recusa de uma afirmação de existência que se refere a uma colectividade somente constituída em abstracto em virtude de um princípio indefinido que se traduz por uma negação ou uma indeterminação: recusa do terceiro excluído e sobretudo das definições não predicativas»[11], No entanto, por muito sedutora que seja, sobretudo pela caracterização da essência da matemática como acto construtivo inscrevendo-se num devir autónomo regrado e efectivo, por muito flexível que seja a sua aplicação (Cavaillès sublinha a comodidade da aplicação da matemática intuicionista à física moderna), a abordagem *intuicionista* deixa demasiadas questões em aberto para que sejam postos de lado os outros pontos de vista lógicos.

Algumas dessas questões são as seguintes. Primeiro, o método intuicionista é ou não irredutível à especificidade do devir matemático? Se sim, que interesse tem para pensar as outras ciências? Se não é, e se se estende à outras ciências, como é que se evidencia a especificidade da cada uma, sobretudo quando Brouwer caracteriza o movimento matemático como ordenamento, e como pensamento racional do mundo («matemática, ciência e linguagem representam as funções principais da actividade da espécie humana, pelas quais esta espécie domina a natureza e mantém nela a ordem», escreve Brouwer em *Mathematik, Wissenschaft und Sprache*). Como é que se distingue do decurso geral da ciência ou até mesmo da cultura? E depois, escreve ainda Cavaillès **«**a independência de uma teoria directa da ciência encontra-se por esta mesma razão gravemente prejudicada**»** já que «a representação de um mundo onde se exerce a actividade

[9] Cavaillès, *La pensée mathématique*, Œuvres Complètes, Paris, Hermann, 1994, p.601.
[10] *Ibid.*
[11] Cavaillès, *Sobre a Lógica*…, §14.

matemática e que determina esta, dando-lhe pelo menos partida e ponto de aplicação, permanece um elemento primordial e talvez dominador»[12] para o pensamento intuicionista.

Os mesmos problemas parecem estar presentes na epistemologia de Brunschvigc. «Em reacção contra o logicismo oriundo de Frege e de Russel na qual é empreendida uma renovação da tradição aristotélica, opõe ao pensamento, enquanto criação que escapa a qualquer norma, a expressão linguística da mesma, que, por ser um fenómeno social, fica sob a alçada tanto das ilusões da cidade como das leis da natureza. A exigência da comunicação leva por um lado ao absoluto aquilo que não era senão um movimento num progresso e tomava o respectivo sentido verdadeiro na passagem que efectuava, por outro lado transpõe-no em sistema cuja estrutura, a forma dedutiva silogística por exemplo, encontra uma justificação numa organização ou numa história sociais que mais não são nelas próprias do que partes da natureza. A *idola tribus* é a revolução do passado, adormecida e petrificada; o pretenso *canon* da ciência, uma regra de pedagogia. Desta forma as querelas em torno do infinito cantoriano só dão indicação de uma resistência a um ímpeto criador novo; são da mesma ordem a reaparição periódica dos mesmos paradoxos, como o de Epiménides, prova da incapacidade que a língua tem em exprimir a potência indeterminada do pensamento»[13]. E, para alem das questões que levanta esta posição de uma imanência racional que escapa a tudo o que não seja ideia, e uma das principais questões que esta posição levanta é a de saber se nesse caso uma filosofia da ciência pode ser outra coisa do que uma «simples explicitação das intenções do sábio» para a qual o progresso dos conteúdos científicos permanece uma exterioridade que a compreensão científica parece não englobar, «nenhuma resposta às questões próprias da epistemologia é dada» e «as ciências todas são confundidas num mesmo movimento»[14]. E, «mais uma vez, escreve Cavaillès, o ser do mundo, de um mundo instituído no exterior e cuja redução em termos de interioridade é uma vocação para a consciência, subsiste como condição determinante da ciência. Conhecer o mundo, compreender o mundo, representa já, como programa, o abandono da autonomia criadora, renúncia a uma

[12] *Ibid*.
[13] *Ibid.* §15.
[14] *Ibid.*

necessidade que a nada se liga a não ser a si própria». Ora, a posição de Bolzano tem precisamente, segundo Cavaillès, a virtude de, simultaneamente, escapar a esta dependência em relação ao mundo e situar a unidade da ciência na lógica inerente à sua estrutura demonstrativa.

O que é que Bolzano nos diz, segundo Cavaillès? «A preocupação predominante de Bolzano é a de acentuar o carácter necessário da ciência. A ciência é antes de mais teoria demonstrada». E, depois: «É possível que pela primeira vez a ciência deixe de ser considerada como simples intermediário entre o espírito humano e o ser em si, dependendo tanto de um como do outro e não tendo realidade própria, mas como um objecto *sui generis*, original quanto à essência, autónomo quanto ao movimento»[15]. Cavaillès subscreve em grande parte a posição de Bolzano e a caracterização que faz do movimento de geração de conceitos (por intermédio das noções de paradigma e de tematização) aparece como um contrapeso dinâmico de uma teoria que, no essencial lhe serve de base. A ciência distingue-se do resto das actividades humanas por ser composta por verdades conceptuais demonstradas. A posição é comum a ambos os pensadores. A ciência «não é nem um absoluto nem um elemento no sistema dos existentes. Se podemos falar da ciência de um país ou de uma época, não se entende nada que esteja incluído nas realidades naturais, mas, no interior do fenómeno sociológico, de afirmações localizadas no espaço e no tempo, o que pertence à ciência. Nem sequer se situa no universo dos objectos de cultura caracterizados pela participação num valor que manifestam. Contrariamente a estes, o modo de actualização que a caracteriza é-lhe extrínseco – não estando ligado de forma assim tão estreita ao valor a ponto tal que seja próprio à sua essência, como no tocante à obra de arte, amalgamá-la à exterioridade acidental de um sistema sensível – por outro lado exige a unidade, quer dizer que não pode conformar-se com uma multiplicidade efectiva de realizações singulares»[16]. A unidade da ciência é sistemática, e estruturada por reenvios mútuos. «Não há diversidade de ciências nem diversos momentos de uma ciência, nem imanência de uma ciência única às disciplinas variadas; estas condicionam-se entre si de forma tal que a significação assim como os

[15] *Ibid.*, §17.
[16] *Ibid*.

resultados de uma exigem, enquanto se trata de uma ciência, a utilização das outras ou a inserção comum num sistema»[17]. É por isso que, segundo Bolzano, mas também segundo Cavaillès que vai sublinhar a importância de ter em conta o aspecto sistemático do universo científico, «uma teoria da ciência não pode ser senão uma teoria da unidade da ciência»[18].

E o que é que caracteriza esta unidade ? «Esta unidade é movimento: como aqui não se trata de um ideal científico mas da ciência realizada, a incompletude e a exigência de progresso fazem parte da definição»[19]. Incompletude, porque o domínio total das verdades em-si, como lembra Bolzano na sua *Wissenschaftslehre*[20] ultrapassa o campo da ciência e exigência de progresso no sentido em que o intuito do trabalho científico é precisamente o de fazer com que os limites da ciência se aproximem cada vez mais dos limites do conjunto das verdades em si. «Só que, escreve Cavaillès, por ser progresso autónomo, dinamismo fechado sobre si próprio, sem começo absoluto nem termo, a ciência move-se fora do tempo – se o tempo significar referência ao que é vivido por uma consciência. Nela, a mudança é acréscimo de volume por geração espontânea de elementos inteligíveis»[21]. J.Sebestik, ao comentar a posição de Cavaillès face a Bolzano em *Sobre a Lógica e a Teoria da Ciência* insiste no entanto no facto de, «na definição bolzaniana da ciência estarem finalmente reunidas – reconciliadas – por um lado a objectividade dos encadeamentos lógicos e por outro lado a respectiva reprodução subjectiva pelo pensamento»[22], já que – podemos acrescentar –, a progressão no domínio das verdades em-si só se pode fazer por sucessivos trabalhos que reactualizam subjectivamente o progresso já feito. Só que, sublinha Cavaillès, «o verdadeiro sentido

[17] Cavaillès, *Sobre a Lógica*, §17
[18] *Ibid.*
[19] *Ibid.*, §18.
[20] Bolzano, *Wissenscahftslehre*, Theory of Science, Edited and translated by Rolf George, University of California Press, Berkeley and Los Angeles, 1972: «*Suppose that all the truths which are now, or ever were, known to any man were somehow collected together, e.g. compiled to the immense domain of truths in themselves, most of which are altogether unknown, this sum is very small […] I take […] the liberty to call an aggregate of truths of a certain kind a science* ».
[21] Cavaillès, *Sobre a Lógica*, §18.
[22] J. Sebestik, *Logique et mathématique chez Bertrand Bolzano*,Paris,Vrin, 1992 p.293.

de uma teoria não está num aspecto que o próprio cientista compreendeu como sendo essencialmente provisório, mas num devir conceptual que não pode parar»[23]. Este devir é caracterizado por uma «absorção do anterior pelo posterior que o justifica e que em certa medida o suprime»[24]. (Vamos reencontrar esta caracterização do devir na apresentação da progressão do pensamento científico por rupturas *paradignáticas*, e abstrações *temáticas*). Por um lado, sublinha Cavaillès, «qualquer conceito ou sistema de conceitos, pelo simples facto de se instituir, é ao mesmo tempo exclusão e exigência do Outro», por outro lado «mesmo no que diz respeito às ciências da natureza, o acréscimo faz-se sem empréstimo ao exterior: existe uma ruptura entre a sensação ou opinião recta e a ciência. A experiência, longe de ser inserção na natureza, é pelo contrário incorporação do mundo ao universo científico». Desde modo, continua, «a autonomia científica é simultaneamente expansão e encerramento: encerramento negativo por recusa de empréstimo ou de resultado exterior. Se o saber total é desprovido de sentido […] o extra-científico radical não o é menos. […] A ciência é um volume *riemaniano* que pode ao mesmo tempo estar fechado e não ter exterior a si próprio»[25].

Perdoem-me que insista na leitura que Cavaillès faz de Bolzano, mas tanto o interesse do trabalho deste como os limites que apresenta a sua filosofia da ciência são essenciais para compreender o campo problemático no qual se insere a posição de Cavaillès. Acto, sentido, movimento paradigmático e temático respondem às insuficiências desta teoria da ciência. Esta, por sua vez, é também uma ciência. Pela seguinte razão. Bolzano escreve que «a soma das regras […] que devemos seguir para dividir o domínio total da verdade em ciências individuais e que devem governar também a redacção dos respectivos tratados merece ser chamado uma ciência. Permito-me chamar-lhe teoria da ciência já que é a ciência que nos ensina a representar as outras ciências»[26]. No entender de Cavaillès, a teoria da ciência assim entendida é parte da ciência por ser ciência da ciência. Os respectivos enunciados não devem no entanto ser constitutivos de movimentos particulares mas devem simplesmente ser uma auto-iluminação do

[23] Cavaillès, *Sobre a Lógica,* §18.
[24] *Ibid.*
[25] *Ibid.*
[26] Bolzano, *Theory of Science*, op.cit., p.2

movimento científico, distinguindo-se deste pela emergência permanente que o caracteriza. A teoria da ciência revela a estrutura que assegura a unidade da ciência. «Ao definir uma estrutura da ciência que não é senão manifestação a si própria do que ela é, os caracteres precedentes adquirem precisão e justificação, não por uma explicitação que teria o seu lugar próprio e seria, por seu turno, objecto de reflexão, mas por uma revelação que não é distinta do revelado, que está presente no seu movimento e é princípio da necessidade que a caracteriza. A estrutura fala sobre si própria. Ora, só há uma forma de se impor por uma autoridade que nada pede ao exterior, só há um modo de afirmação incondicional: a demonstração. A estrutura da ciência, não só é demonstração mas confunde-se com a demonstração. Nela se encontram realmente os traços essenciais: unidade, progressão necessária e indefinida, por fim fecho sobre si própria»[27]. E acrescenta : «a demonstração, pelo simples facto de instituir o fim, estende e ramifica o domínio criado por intermédio de combinações que estabelece assim que são possíveis. Por fim não pode aliar-se ao não demonstrado»[28] E escreve ainda : «A ciência, sendo, é inteiramente demonstração, quer dizer lógica»[29].

Só que, esta estrutura que funciona, na leitura de Cavaillès, como *verum index sui*, levanta um problema que é o «de apreender este princípio no seu movimento gerador, de reencontrar esta estrutura não por descrição mas de forma apodíctica enquanto se desenrola e se demonstra a si própria. Por outras palavras, a teoria da ciência é um *a priori*, que não é anterior à ciência mas que é alma da ciência, não tendo requisitos exteriores mas exigindo por seu turno a ciência»[30]. Desta forma, se a teoria bolzaniana: «evita a subordinação a um existente histórico ou ao absoluto da consciência, tem de ser ela própria a instituir a totalidade daquilo que atinge, distinguindo de seguida, caso possa, o elemento essencial permanente daquilo que por ele é móvel. Construção de uma teoria pura dos encadeamentos racionais por um lado, por outro lado relação à ciência desenvolvida»[31]. No entender de Cavaillès a epistemologia científica não pode constituir directamente uma teoria pura dos encadeamentos

[27] Cavaillès, *Sobre a Lógica,* §19.
[28] *Ibid.*
[29] *Ibid*.
[30] *Ibid.*, §20.
[31] *Ibid.*

racionais sem uma analítica que permite obter o conteúdo dos objectos desses encadeamentos nem sem uma ontologia que perfaz o objecto como *ser*. Assim se inicia o trecho de *Sobre a lógica e a teoria da ciência* em que Cavaillès desenvolve as noções de movimento longitudinal da estrutura demonstrativa, ou paradigma, e movimento vertical, ou temático. É interessante, mais uma vez, esclarecer esta posição com aquela que está expressa na conferencia sobre *O pensamento matemático* onde podemos ler que «o conhecimento matemático [nos] impede […] de admitir objectos como existindo independentemente do sistema que se cumpre sobre esses objectos e até independentemente de um encadeamento necessário [...]. Somos obrigados a dizer de cada vez: é de correlatos de uma actividade que se trata. Aquilo que neles pensamos não são senão as regras de raciocínio matemático que são exigidas pelos problemas que se põem»[32]. Mas voltemos ainda às considerações que Cavaillès tece sobre Bolzano.

«Eis o que acarreta uma teoria fiel da ciência: o demonstrável não é tido em consideração. Mas unidade não significa uniformidade: sob a identidade de essência o demonstrativo pode revestir aspectos múltiplos. «É possível, em particular, situar frente a frente uma zona de singularidades na qual o demonstrável adere à demonstração a ponto de ser indiscernível e caracteriza um momento único da ciência, e por outro lado, o conjunto dos encadeamentos propriamente ditos que, apesar de não se imporem senão na medida em que efectivamente demonstram qualquer coisa e não serem por conseguinte renováveis na integra, apresentam por grupos um parentesco quanto ao tipo, que é marca da unidade de movimentos e é manifestável em abstracto»[33]. Inseparáveis dos objectos e da singularidade dos seus conteúdos enquanto demonstram, os encadeamentos demonstrativos são no entanto pensáveis numa unidade abstracta a partir da variação dos conteúdos singulares sobre os quais incidem, o com os quais decorrem. «O processo de separação é duplo: longitudinal, ou co-extensivo ao encadeamento demonstrativo, vertical ou instaurando um sistema de ligações novo no qual o antigo tem o papel de base de partida, e já não o de estádio atravessado por um movimento, mas sim o de objecto de reflexão no aspecto que actualmente tem. Em cada

[32] *La pensée mathématique*, Œuvres Complètes, op.cit. p.601.
[33] Cavaillès, *Sobre a Lógica*, § 21.

caso manifesta-se uma propriedade constitutiva da essência do pensamento: o paradigma e o temático»[34].

E o que é o paradigma no vocabulário de Cavaillès? «O paradigma é característico da actualização. Não de uma actualização hic et nunc no vivido cujo extrínseco é eliminado, na simples apropriação do pensamento por si próprio; mas de uma actualização exigida pelo sentido do que é instituído : uma relação que enquanto tal não se afirma senão na singularidade de realização do encadeamento, mas que não requer esta senão enquanto singularidade qualquer, portanto ao instituí-la está a suprimi-la e revela deste modo um princípio interno de variação»[35]. A *tematização*, por seu turno, «parte do encadeamento que desta vez é apreendido no seu voo, que é trajectória que se muda em sentido. O pensamento já se não dirige para o termo criado, antes parte da forma de criar para determinar o princípio desta por intermédio de uma abstracção igual em natureza à outra, mas transversalmente dirigida»[36]. A *tematização* responde à necessidade de realização da formalização pela sobreposição sistematizada das regras que regem as estruturas ao desenho destas últimas.

Mas voltemos ao paradigma. «Do raciocínio aritmético sobre o número inteiro finito aos encadeamentos mais abstractos, escreve Cavaillès, opera-se a mesma dissociação libertadora de sentido»[37]. Gilles-Gaston Granger, no terceiro capítulo de *Pour la Connaissance Philosophique*, dá um exemplo do processo paradigmático extraído da história da geometria. Os antigos definiam as três espécies de cónicas (elipse, hipérbole, parábola) como resultado de secções «por planos perpendiculares a uma geratriz […] de cones de ângulo com o vértice agudo, obtuso ou recto»[38]. É partir de um cone particular, definido segundo um tipo de ângulo do respectivo vértice e de um tipo de operação idêntica (o corte por um plano perpendicular a uma geratriz) que se obtém uma cónica particular. Passa-se ao *paradigma* da operação a partir do momento em que, como Apollonius, se mostra «que as três cónicas podem ser geradas num mesmo cone fazendo

[34] *Ibid.*
[35] *Ibid*. §22.
[36] *Ibid.* §25.
[37] *Ibid.*
[38] G-G. Granger, *Pour la Connaissance Philosophique*, Paris, Odile Jacob, 1988, p.71.

variar o plano de secção», qualquer que seja o cone. O que é importante, desde logo, não é já a especificidade das configurações que servem de ponto de apoio à operação por intermédio da qual se obtém uma das cónicas, o que importa é a configuração *enquanto podendo variar* relacionada com a operação cujo sentido assim se destaca. O objecto matemático e científico é a *operação posta a nu* que é correlativa a uma diversidade de aplicações. A unidade paradigmática obtida é, no entanto, dotada de um princípio de variação da operação que não existia à partida. Este princípio é o *sentido* da operação, «o ser da relação». Como ser da relação, escreve Cavaillès, «deve entender-se o que esta acrescenta à sua origem, diferente pois da necessidade que a torna una, afirmando assim apesar dessa necessidade uma independência que se traduz em indiferença relativa, que é fonte de pluralidade»[39]. O movimento longitudinal paradigmático é assim «uma dissociação libertadora de sentido». Cavaillès comenta esta afirmação com o fragmento 142 do poema de Parménides: «Se é se diz do *um* que é e *um* do ser um e se *ser* e *um* não são a mesma coisa mas pertencem igualmente a essa coisa… não se seguirá necessariamente que o um que é é um todo e que o um e o ser dele são as partes?». Por este comentário parece chamar a nossa atenção para o facto de a totalidade una (a da operação obtida por variação dos argumentos) ter partes que conservam o um e o todo, podendo o um (do sentido) ser uma pluralidade infinita, como a unidade da operação está em correlação com uma infinidade de conteúdos. «O *um que é* será uma pluralidade infinita» acrescenta, citando agora o fragmento 143.

Cavaillès caracterizaria da mesma forma o ser da relação que a operação introduz. O exemplo que escolhe é o da relação de identidade (A=A) no qual o que importa não é a singularidade repetida – singularidade da condição na qual a relação pode ser estabelecida – mas a operação que é posta em evidência. A dissociação do repetido e da repetição é, na verdade, dissociação que liberta o sentido da operação da singularidade inicial na qual encontrava apoio e com a qual operava. Cada promoção de objectos em matemática resulta de uma passagem da *ligação-acto* à *ligação-tipo* «como na fuga indefinida rumo ao sentido que, extraído graças ao cenário inteligível e fabricado com este, de imediato se escapa «como num sonho» e

[39] Cavaillès, *Sobre a Lógica,* § 22.

institui dessa forma a sua originalidade»[40]*,* escreve Cavaillès. É a singularidade que serve de ponto de apoio ao *acto* matemático, mas é o *sentido* a que dá origem que constitui o *tipo* de relação que faz a originalidade de uma operação. Em «matemática*,* escreve G.G. Granger a propósito do movimento paradigmático, a efectuação do acto extrai o necessário e o suficiente das condições que lhe são próprias, abrindo assim caminho à infinita renovação das variantes autorizadas no campo que estas condições delimitam»[41]. «Há polimorfia na unidade do encadeamento racional»[42] afirma Cavaillès. Os objectos da experiência matemática são estas operações tornadas necessárias que resultam «[d]as sucessivas fracturas de independência que [...] destacam do que é anterior o perfil imperioso do que vem a seguir necessariamente e para o ultrapassar»[43], o que lembra o que Cavaillès escreveu acerca de Bolzano. «O que aqui tem importância é o desprendimento que se opera cada vez que se suprime uma singularidade: o que no cálculo lógico é representado pela regra de substituição, a possibilidade, no elemento novo, de substituir por outro qualquer aquele do qual de facto procede, equivalente a este do novo ponto de vista atingido»[44]. «Assim se explica, a cada uma das etapas, a distinção entre matéria, ou singularidade de origem, e forma, ou sentido actual»[45]. É importante, no entanto, não cair na armadilha de um formalismo que projectaria «no absoluto, uma distinção que teria lugar de uma ponta à outra do encadeamento geral, como uma falha contínua atravessando os nós das distinções particulares» é essa por exemplo, na leitura de Cavaillès, a posição errada de Kant no que diz respeito ao pensamento da matemática. Um formalismo está sempre ligado aos conteúdos dos actos de que é oriundo e que prolonga. O acto ou operação efectiva pode distinguir-se do sentido sem que se negligencie «o motor de passagem necessariamente progressiva do acto ao seu sentido: não há sentido sem acto, nem acto novo sem o

[40] *Ibid.*

[41] G.-G. Granger, *Pour la Connaissance Philosophique*, p.73 [«[...] [En] mathématique, l' effectuation de l' acte dégage le nécessaire et le suffisant de ses propres conditions, ouvrant ainsi carrière à l' infini renouvellement des variations autorisées dans le champ que ces conditions délimitent»]

[42] Cavaillès, *Sobre a Lógica*, §22.

[43] *Ibid.*

[44] Cavaillès, *Sobre a Lógica,* § 23.

[45] *Ibid.* §24.

sentido que o gera»[46] A ilusão da separação entre forma e conteúdo, no que diz respeito à matemática (e talvez à apresentação axiomática das teorias) é resultado da apresentação artificial e «da exposição sistemática de uma teoria matemática, na qual o recenseamento inicial dos procedimentos é oriundo tanto da análise do sentido derradeiro dos encadeamentos como do que há de artificial naquilo que consiste em considerar estes como preparando-o sem sobressalto intermédio ou ulterior. Trata-se do *momento da variável*: ao substituir as determinações de actos pelo lugar vazio para uma substituição, elevamo-nos a um grau de abstracção que dá a ilusão da existência de um formal irredutível»[47].

Pelo contrário, insiste Cavaillès, «a formalização não se realiza senão quando ao desenho das estruturas se sobrepõem de forma sistematizada as regras que as regem»[48] É desta forma que Cavaillès inicia em *Sobre a Lógica e a Teoria da Ciência,* a apresentação do movimento vertical temático característico da abstracção conceptual próprio às ciências demonstrativas. O que é que se entende por este termo ? Em *Méthode Axiomatique et Formalisme* (uma das teses de doutoramento apresentadas em 1937 por J. Cavaillès) a *tematização* é apresentada como sendo a «transformação de uma operação num elemento de um campo operatório superior»[49]. Esta transformação efectua-se por uma reflexão que tem por objecto o antigo sistema de ligação (operação de grau inferior, obtida pelo movimento paradigmático) que se torna base de partida «*e já não estádio atravessado por um movimento*»[50] do qual o paradigma dá o sentido. A tematização é o movimento vertical «que [se faz] instaurando um sistema de ligações novo»[51]. Consiste em tomar como ponto de partida o encadeamento apreendido como trajectória que se transforma em sentido; a tematização dá o princípio deste, ao partir da forma de criar. Já não se trata de dar conta da criação da forma, mas do "como" desta. Segundo Cavaillès, encontram-se exemplos no desenvolvimento da matemática: «teoria dos grupos, teoria das operações lineares, das

[46] *Ibid.* §24.
[47] *Ibid.*
[48] *Ibid.*, §25.
[49] Cavaillès, *Méthode Axiomatique et Formalisme*, p.177, in *Œuvres Complètes*, op. cit. p.185.
[50] Cavaillès, *Sobre a Lógica*, §24
[51] *Ibid*.

matrizes [e] topologia das transformações topológicas»[52]. No que diz respeito à teoria dos grupos, a tematização consiste em abstrair uma forma de forma mobilizando e tornando variável a própria forma obtida por um processo paradigmático. Tome-se o exemplo seguinte da adição. Primeiro é uma operação sobre números inteiros; o processo longitudinal suprime a singularidade dos argumentos – os números inteiros tornam-se elementos quaisquer que se podem fazer variar: podemos efectuar a operação com vectores, com matrizes, diz Cavaillès. Assim sendo, o sentido da adição não depende já da especificidade dos argumentos. O processo vertical da tematização deve levar à apreensão dos princípios formais da adição enquanto esta é uma de várias leis de composição. Permite extrair os axiomas da teoria dos grupos comutativos. Desta forma, o grupo **G** da operação " * " será definido pela propriedade que enuncia a existência de um *produto interno*; pela propriedade de *associatividade*; pela existência de um *elemento neutro* do conjunto das constantes que satisfazem a operação e pela existência de um elemento chamado *inverso* de tal modo que um qualquer elemento *a* pelo seu inverso dê o elemento neutro ($a * a^{-1} = e$). O processo vertical permite atingir um grau superior que toma o anterior como base de partida e objecto de reflexão. Do esforço temático para a obtenção de uma teoria efectiva das operações nasceram historicamente os cálculos de Boole, Schröder e Frege, assim como os sistemas formais gerais, segundo Cavaillès.

O processo que gera um desenvolvimento inteligível novo encontra-se condicionado pelo «carácter estruturado dos encadeamentos cuja lei morfológica tem estatuto de princípio». Mas encontra-se igualmente condicionado pela «dissociação entre sentido e acto da qual o geral resulta longitudinalmente», podemos ler. É o que permite retomar a distinção anterior, entre acto e sentido, e introduzir uma distinção nova, entre *sentido instituinte* e *sentido instituído*. «No momento em que se liberta o sentido aparece uma dualidade entre sentido *instituinte* e sentido *instituído*, entre a significação de uma operação enquanto é *operada* – sem ter em consideração o problema cuja exigência criou a singularidade activa da mesma, e despojada da partida pela substituição possível do acidental do início – e a significação da mesma enquanto *operante*, na qual se entende o

[52] Cavaillès, *Sobre a Lógica*, §25.

sentido já não da passagem mas do *quomodo* desta, onde o resultado já não se considera na essência, por muito pura que tenha passado a ser, mas na simples posição de um resultado que resulta, por conseguinte na indicação dos modos que o geraram e sobre o qual recai o interesse»[53]. Como exemplo, Cavaillès toma a adição. O paradigma da operação aditiva – o sentido instituído desta – é insensível à natureza dos objectos sobre os quais se efectua – e deste modo torna-se «*multiplicação ou adição abstracta*», por assim dizer, purificada. Mas, esta forma desnudada não deixa de servir de ponto de partida para o processo que deve concluir-se pela apreensão dos princípios da *operação operada*, que são os elementos constitutivos da *operação operante* que aparece num plano novo. A tematização extrai o sentido instituinte da lei de composição que é edificada pelo conjunto de regras de associatividade, de comutatividade e de combinação distributiva com a operação de multiplicação. Há solidariedade entre o instituinte e o instituído, esta traduz-se pelo reenvio mútuo do *sentido instituinte* ao *sentido instituído* ou, como escreve Cavaillès, os dois processos entrelaçam-se mutuamente na medida em que «são oriundos da mesma surreição de sentido»[54] e podem dar lugar a uma «complicação inteligível ilimitada, que mostra o quão simplista era a imagem da separação entre matéria e forma num encadeamento racional e como passa a ser indispensável um estudo sistemático das formas»[55].

Se não há formalismo sem sintaxe nem sintaxe sem outro formalismo que a desenvolva, se tanto a forma necessária como o princípio se encontram iluminados, formalizar é fundar e a teoria da ciência passa a herdar o legado da lógica e talvez até a ultrapassasse na medida em que se confundiria com o sistema de todos os formalismos possíveis e absorveria a totalidade das demonstrações. Cavaillès prossegue com a análise e a discussão do esforços de formalização de Frege, Carnap e Tarski na medida em que parecem opor-se ao devir do conhecimento demonstrativo (a progressivo faz parte da essência) e à criatividade estruturada que o caracteriza e cuja trama dinâmica é a dos elos internos entre as ideias.

[53] *Ibid*.
[54] *Ibid*., §26.
[55] *Ibid*.

EL MITO DEL MÉTODO Y LAS ESTRATÉGIAS DEL DESCUBRIMIENTO CIENTÍFICO. INDUCCIÓN, ABDUCCIÓN, PREDUCCIÓN

Andrés Rivadulla
Dpto. de Lógica y Filosofía de la Ciencia
Universidad Complutense

1. Introducción

En un sentido filosóficamente neutral de *ciencia*, el término *método científico* supone la existencia de unos procedimientos cuya aplicación garantiza el logro de *conocimiento.* De hecho, para Karl Popper, una tarea de la filosofía de la ciencia es ofrecer un análisis lógico del método científico. Esta es la reflexión que aparece en las primeras seis líneas de su *Lógica de la Investigación científica* de 1935.

Particular importancia para la teoría de la ciencia ha tenido el compromiso con un cierto *fetichismo* del método en algunos momentos de la cultura de Occidente. Por ejemplo

1) La pretensión en el siglo XIX de Comte, Quetelet y Laplace de extender el método de la *mecánica celeste* a todas las ciencias, inclusive las ciencias humanas y sociales, que deberían acabar convirtiéndose en una especie de *mecánica* o *física social*.

2) La propuesta de Karl Popper en pleno siglo XX de identificar la metodología *falsacionista* con el *método científico.* En efecto, para Popper (1966: 260)

> The method of science is rather to look out for facts which may refute the theory. This is what we call testing a theory (...) This view of scientific method is corroborated by the history of science, which shows that scientific theories are often overthrown by experiments, and that the overthrow of theories is indeed the vehicle of scientific progress.

El *fetichismo del método* obedece a que se tiene una creencia estricta acerca de en qué consiste la ciencia, y se diseña el/los métodos que más eficazmente lleva(n) al fin, el *conocimiento*. La cuestión es, naturalmente, si esta creencia está justificada, si hay realmente un método propio de la ciencia.

La creencia en la existencia del método científico, y sobre todo en su versión más radical que asume un *fetichismo del método*, tuvo al menos dos consecuencias graves y divergentes:

> **1.** Una reacción conformista en el ámbito de la filosofía, que tuvo como consecuencia la separación de metas entre ciencias humanas y sociales por una parte -cuyo objetivo sería la *comprensión*-, y ciencias de la Naturaleza, por otra -cuyo objetivo debería ser la *explicación*. Esta reacción cristalizó en la aceptación de la separación metodológica entre humanidades (*Geisteswissenschaften*) y ciencias de la Naturaleza (*Naturwissenschaften*). Y supuso, por una parte, romper con una tradición bimilenaria de *cultura* occidental *unificada*, y por otra parte, ceder el monopolio de la *racionalidad* a las ciencias de la Naturaleza (Rivadulla: 2008).
> **2.** Una reacción reivindicativa del *contexto de descubrimiento,* que, sin renunciar al control empírico de las hipótesis, enfatiza la importancia de los aspectos creativos de la ciencia. Esto supuso pensar la ciencia como parte de la cultura, y romper con el fetichismo del método, intuyendo la complejidad de objetivos y procedimientos de la práctica científica en su totalidad.

2. Dudas en torno al método científico, y consecuencias del *fetichismo* del método

La idea de la existencia de un método científico no ha encontrado siempre el apoyo unánime de los filósofos de la ciencia, a veces incluso de aquellos que por su orientación analítica parecerían mejor inclinados a asumirlo sin discusión. Es por ejemplo el caso de Bertrand Russell, quien en su libro *La perspectiva científica*, Cap. III, reconoce la existencia de tres tipos de *limitaciones del método científico*: 1) La duda respecto a la validez de la inducción: "la inducción sigue siendo un problema de lógica no resuelto" (p. 64); 2)

La dificultad de inferir de lo que ha sido experimentado a lo que no lo ha sido: "¿hay siempre circunstancias tales que nos permitan inferir de una serie de hechos conocidos, que algún otro hecho ha ocurrido, está ocurriendo u ocurrirá?" (p.68); y 3) El hecho de que tal inferencia puede ser de un carácter extremadamente abstracto, y proporcione, por consiguiente, menos información de la que resulta cuando se emplea el lenguaje ordinario.

Por su parte el post-positivista Ernest Nagel mantuvo (1961: 12-13) que sería una mala interpretación afirmar

> that the practice of scientific method consists in following prescribed rules for making experimental discoveries or for finding satisfactory explanations for matters of established fact. *There are no rules of discovery and invention in science any more than there are such rules in the arts.*

O sea, que no se puede mantener que

> the practice of scientific method consists in the use in all inquiries of some special set of techniques (...), irrespective of the subject matter or the problem under investigation.

La posición completamente opuesta es la que protagoniza Karl Popper (1974), en quien el *fetichismo del método* repercute con toda crudeza en su convicción de que "*el darwinismo no es una teoría científica, porque no es testable*."

Si esta opinión de Popper nos repugna de entrada, es porque asumimos una concepción *intuitiva* de *ciencia*, en la que no parece aceptable el rechazo como no científicas de prácticas culturales tales como la biología evolucionista, y otras similares: paleontología, geofísica, y en general las *ciencias observacionales de la Naturaleza*. Esto es lo que opinan filósofos contemporáneos tales como Hilary Putnam (1981: 198-199):

> we accept the Darwinian theory of evolution by natural selection as what Peirce called an 'abduction' , or what has recently been called an inference to the best explanation", "because it provides a *plausible explanation* of an enormous amount of data."

Y Paul R. Thagard (1978:77):

> Inference to the best explanation is common in the history of science. An explicit example of an argument to the best explanation is Charles Darwin's long argument for his theory of the evolution of species by means of natural selection. In his book *The Origin of Species* he cites a large array of facts which are explained by the theory of evolution but which are inexplicable on the then-accepted view that species were independently created by God.

Para estos filósofos es la *abducción*, en cuanto procedimiento aceptado para la propuesta de hipótesis científicas, la que le da carácter científico a la teoría darwinista.

El rechazo de Popper del darwinismo como ciencia es paralelo a su menosprecio del *contexto de descubrimiento* (1959: 31). Este descuido de Popper por el contexto de descubrimiento empobreció su filosofía de la ciencia hasta el punto de ignorar formas fértiles de la creatividad científica como la *analogía*, o incluso la *serendipia* (la *chiripa*). Por no hablar de estrategias clásicas del descubrimiento: la inducción, la abducción y la preducción. Es verdad que Popper asume que la ciencia empieza siempre con hipótesis. Pero no se percata de que la forma cómo estas hipótesis llegan a la ciencia es también metodológicamente relevante. El caso de la serendipia es elocuente.

La *serendipia* (descubrimiento casual de algo), quizás mejor la *pseudoserendipia* (Royston M. Roberts, 2004): descubrimiento casual que culmina un camino de búsqueda, es una frecuente forma muy fértil de descubrimiento. Ejemplos típicos de serendipia serían:

- El Principio de Arquímedes
- La postulación por August Kekulé en 1865 de la estructura hexagonal del benceno
- El descubrimiento de los Rayos X por Wilhelm Conrad Roentgen en 1895
- El descubrimiento de la radiactividad natural por Henri Becquerel en 1896
- El descubrimiento de la penicilina por Alexander Fleming en 1922

- El descubrimiento de la estructura de doble hélice de la molécula de ADN por James Watson y Francis Crick, etc.

Otro caso genuino de serendipia en geofísica es el del descubrimiento de las bandas magnéticas en los fondos oceánicos, de gran relevancia para la consolidación de la teoría de la dinámica de placas tectónicas. Ron Mason (2001: 43) lo refiere del modo siguiente:

> The discovery of sea floor magnetic stripes was *serendipitious*: we were not looking for them, nor could we have been, because no one knew they existed! […] The recognition of sea floor magnetic stripes led immediately to the hypothesis of large displacements of the ocean floor which,… stimulated interest in the mobility of the oceanic crust, and it raised interesting questions. (…) For most people these results placed beyond doubt the existence of ocean floor spreading, a factor fundamental to the concept of plate tectonics.

Junto a la serendipia, el *razonamiento por analogía* constituye un vehículo importante para la creatividad científica. Se le ha considerado ampliamente como una forma de inducción: si ciertos individuos tienen determinadas propiedades, y observamos en uno nuevo la mayor parte de estas propiedades, concluimos que las tiene todas. También se ha destacado el papel de la analogía para la construcción de metáforas científicas (Cfr. Rivadulla 2004:149 y 2006). Pero yo quiero insistir aquí en su fertilidad heurística para encauzar el pensamiento científico en determinadas direcciones, e. d. como forma fecunda de creatividad. Vaya como ejemplo la propuesta del modelo nuclear de capas por analogía con el modelo atómico de capas. Robert Eisberg y Robert Resnick, citados por Rivadulla (2004: 149-150), lo expresan del modo siguiente:

> La analogía entre números mágicos nucleares y atómicos indujo a muchos a buscar una explicación de los fenómenos nucleares similar a la explicación del fenómeno atómico.

La existencia de estas dos herramientas heurísticas, *serendipia* y *analogía*, sugiere que las formas de creatividad merecen ser tomadas seriamente en cuenta por la metodología de la ciencia. En ellas el

componente psicológico es fuerte, pero contra Popper (1959: 31-32) para quien

> the question how it happens that a new idea occurs to a man - whether it is a musical theme, a dramatic conflict, or a scientific theory- may be of great interest to empirical psychology; but it is irrelevant to the logical analysis of scientific knowledge

parece razonable afirmar que, para la metodología de la ciencia, el contexto de descubrimiento es tan importante como el contexto de justificación.

En lo que sigue me voy a referir a aquellas formas de creatividad en que el razonamiento científico está más próximo a la lógica que a la psicología: la *inducción*, la *abducción* y la *preducción*. Tal variedad muestra la *pluralidad metodológica* en ciencia.

3. La inducción

Es un hecho que Charles Peirce y Rudolf Carnap sitúan sin ambigüedad a la inducción en el contexto de justificación. En efecto, para Peirce

> La inducción es la comprobación experimental de una teoría. (…) Parte de una teoría y mide su grado de acuerdo con los hechos. (CP, 5.145)

> La inducción consiste en partir de una teoría, deducir de ella predicciones de fenómenos, y observar estos fenómenos, a fin de ver *cuánto* concuerdan con la teoría. (CP, 5.170)

Se trata desde luego de un anticipo de la *lógica inductiva* de Carnap.

No obstante, desde Aristóteles (Rivadulla 2007: 200-201) la inducción se entiende fundamentalmente como una *inferencia conservadora de la verdad y ampliadora del contenido.* En sentido aristotélico la *generalización inductiva* es la forma lógica, no psicológica, de *descubrimiento* de los principios y leyes generales de la ciencia, de la verdad. Por ello:

1. fue criticada, y renovada, por Francis Bacon en el siglo XVII
2. fue declarada *inexistente*, así como *imposible* la probabilidad inductiva, por Karl Popper en el siglo XX.

Ahora bien, en relación a este segundo punto, no se puede olvidar que el *antiinductivismo de teorías* había sido anticipado en el siglo XX tanto por Pierre Duhem (1906: 327 y 334): "Es imposible construir una teoría utilizando un método puramente inductivo", como por Albert Einstein (1936: 276): "No existe un método inductivo que nos conduzca a los conceptos fundamentales de la física."

La razonabilidad *Humeano-Popperiana* del *antiinductivismo de teorías* reside en la *invalidez lógica* de la inducción como método para el descubrimiento de verdades.

Aún así, Hans Reichenbach (1936: 4) pensaba que, si la meta de la investigación científica, consistente en el establecimiento de leyes de los eventos naturales, es alcanzable, lo será por medio del método inductivo. Opinión que todavía mantenía en (1938: 383): "The methods of induction ... always will remain the genuine methods of scientific discovery."

Afortunadamente el debate en torno a la inducción no se ha centrado sólo en la disputa entre posiciones extremas. También ha habido posturas intermedias, más pragmáticas, en torno a la aplicación metodológica de la inducción. Este es el caso, por ejemplo, de Isaac Newton, quien en la *Regula Philosophandi* IV de su *Principios Matemáticos de la Física*, 1687, mantiene que

> las proposiciones obtenidas por inducción a partir de los fenómenos han de ser tenidas, pese a las hipótesis contrarias, por verdaderas o muy aproximadas, hasta que otros fenómenos las hagan más exactas o sujetas a excepciones.

Así como en su *Óptica*, 1706, Libro III, Parte I, donde afirma que

> aunque los argumentos a partir de observaciones y experimentos por inducción no constituyan una demostración de las conclusiones generales, *con todo, es el mejor modo de argumentar que admite la naturaleza de las cosas* (...) Si de los fenómenos no surge ninguna excepción, las conclusiones pueden proclamarse en general. Pero, si ... surgiese alguna

> excepción de los experimentos, habrán de comenzar a proclamarse con las excepciones pertinentes.

Es claro pues que Newton mantiene una posición pragmática respecto de las posibilidades de la inducción: las generalizaciones inductivas pueden tener excepciones. Esta forma particular de inducción es, a falta de otra mejor forma de razonamiento, la que, según él, proporciona el medio para llevar a cabo descubrimientos científicos.

Este pragmatismo inductivo lo encontramos de nuevo entre los miembros del primitivo Círculo de Viena, en cuyo documento programático, redactado por Carnap, Hahn y Neurath (1929: 24), se dice

> El método de *inducción* se puede aplicar, con suficiente fundamento o sin él, allí donde conduzca a resultados fructíferos. La seguridad no la garantiza. Pero la sensatez epistemológica requiere que a una inferencia inductiva sólo se le conceda importancia si puede ser comprobada empíricamente.

De modo parecido, Bertrand Russell (1949: 57), aunque concibe que "Todas las leyes científicas descansan sobre la inducción", mantiene no obstante que la *inducción por enumeración simple* "es una manera muy peligrosa de argumentar", por lo que reconoce que "considerada como un proceso lógico [la inducción] está abierta a la duda, y no es capaz de dar certeza." En esta crítica coincide plenamente con Francis Bacon.

Finalmente, es curioso constatar que la forma del argumento inductivo para Russell es claramente una *abducción* (basta reordenar las premisas):

- Si cierta hipótesis es verdadera, tales y cuales hechos serán observables;
- Ahora bien, estos hechos son observados;
- Consiguientemente, la hipótesis es probablemente verdadera.

Estas reflexiones permiten extraer las siguientes conclusiones respecto de la inducción

> **1.** Considerada estrictamente como una inferencia conservadora de la verdad y ampliadora del contenido, la inducción es lógicamente ilegítima, no es un método válido para el descubrimiento científico, *pace* Popper.
> **2.** Considerada en sentido práctico, como *heurística para la propuesta de hipótesis*, que el control empírico se encargará de aceptar (al menos provisoriamente) o rechazar, la inducción es aceptable. Es una *estrategia más del descubrimiento científico.*
> **3.** Su indefinición lógica contribuye a su confusión con la *abducción*.

4. Abducción e inferencia a la explicación más probablemente verdadera

Charles S. Peirce, (CP, 5.171) concebía la abducción como "el proceso de formación de una hipótesis explicativa. Es la única operación lógica que introduce una nueva idea". El lugar de la abducción es el contexto de descubrimiento: "La abducción consiste en estudiar hechos e inventar una teoría que los explique" (CP, 5.145). La forma archiconocida del argumento abductivo es, según Peirce (CP, 5.189), la siguiente:

- El hecho sorprendente, *C*, es observado;
- Pero si *A* fuera verdadera, por supuesto se daría *C*;
- Luego hay razón para sospechar que *A* es verdadera.

La abducción es una forma de razonamiento, propia -pero no exclusiva- de las *ciencias observacionales de la naturaleza*, por medio de la cual se selecciona tentativamente como la más razonable aquella hipótesis, de entre las que compiten entre sí, que, a criterio del investigador, mejor compatibilidad muestra con los datos disponibles. O, como dice John R. Josephson (1994, p. 5), siguiendo la opinión generalmente aceptada

> *La abducción, o inferencia a la mejor explicación,* es una forma de inferencia que va de los datos que describen algo a una hipótesis que mejor explica o da cuenta de los datos.

El razonamiento abductivo seguiría el siguiente esquema, según Josephson, que como podemos ver se asemeja mucho a la forma que toma la inducción para Russell:

- *D* es una colección de datos (hechos, observaciones, etc).
- *H* explica *D* (si fuera verdadera, explicaría *D*).
- Ninguna otra hipótesis puede explicar *D* tan bien como lo hace *H*.
- Por tanto, *H* es *probablemente* verdadera.

Y esta forma de argumentar nos lleva a plantearnos (Cfr. Rivadulla 2004: 48-59, y 2007:205-207) si es concebible entonces un *enfoque bayesiano* de la inferencia abductiva.

Como sabemos, el problema de la inferencia bayesiana reside en que si, tras una acumulación de evidencia favorable a una hipótesis -que haría concebir la esperanza de que su probabilidad de ser verdadera está próxima a 1- una nueva evidencia la contradijese, entonces una hipótesis harto probablemente verdadera quedaría refutada.

Pero veamos cómo una *ficción histórica* puede ayudarnos a comprender los límites del razonamiento bayesiano. Supongamos que la *Cuestión I* que Isaac Newton incluye en su *Optics*, 1704

> ¿Acaso no actúan a distancia los cuerpos sobre la luz, y por medio de esta acción doblan sus rayos; y no es esta acción (*coeteris paribus*) tanto más fuerte cuanto menor es la distancia?

hubiese sido contestada antes, digamos en tiempos de Newton mismo. Como en *aproximación newtoniana* (Cfr. Rivadulla 2003: 219-221) la desviación calculada es de 0,87 segundos de arco, entonces, a la vista del fracaso de su predicción, desmentida por la observación de que la desviación de la luz es realmente de 1,75 segundos de arco, Newton mismo se habría visto forzado a reconocer la falsedad de su teoría, independientemente de posteriores confirmaciones. Ningún enfoque

bayesiano podría haberla socorrido, y ¡la historia de la física habría sido diferente!

Por eso me permito recomendar (Rivadulla 2004: 13), que "haríamos bien en desterrar la pregunta acerca de qué probabilidad de ser verdadera tiene determinada teoría científica, o en qué medida una teoría es más probablemente verdadera que otra." Pero no porque el Teorema de Bayes sea incorrecto, sino porque no prescribe condiciones de selección de predicciones ni su secuencia temporal. En el bien entendido que el rechazo del enfoque bayesiano de la abducción no arrastra consigo en su caída a la propia abducción.

5. Abducción falible en las ciencias naturales observacionales. El caso de la geofísica

El geólogo austríaco Eduard Suess (1831-1914) había observado que la glossopteris, una forma primitiva de helecho, se encontraba en registro fósil en Sudamérica, África y la India, perteneciente al carbonífero y el pérmico, las dos etapas finales de la Era Primaria de la Tierra, así como en el Triásico, la etapa inicial de la Era Secundaria. Su *explicación abductiva* fue que los continentes actuales debieron estar unidos hace unos 300 millones de años formando un supercontinente, al que bautizó como Gondwánaland.

La hipótesis de Gondwána se insertaba en su *modelo de Tierra en contracción*, o de *granada seca* (drying apple model), con el que Suess en *La Faz de la Tierra*, 1883-1909, pretendía dar cuenta de la formación de montañas, continentes y océanos. Este modelo conjeturaba los procesos sucesivos siguientes:

- El enfriamiento del material fundido del interior de la tierra produciría la contracción de la corteza y el colapso de partes de ella, que el océano acabaría sumergiendo.
- Fuerzas laterales de la partes no colapsadas sobre las colapsadas producirían la formación de montañas.
- Con el continuo enfriamiento, sucesivos colapsos de la corteza producirían intercambios de tierras y mares, lo que explicaría la presencia de fósiles marinos en tierras emergidas.

La debilidad del modelo de contracción de Suess la pondrían de manifiesto tres hechos, según Alfred Wegener (1929):

> **1.** el descubrimiento de los mantos de corrimiento de los Alpes (p.21)
> **2.** el descubrimiento del radio, cuya desintegración generaría calor continuamente en el interior de la Tierra (p. 22)
> **3.** la *isostasia* y el *rebote fino-escandinavo* (pp. 23-24 de la edición española)

La conclusión provisional de Wegener fue: "La suposición,…, de que la situación relativa de los bloques continentales no ha cambiado (…) debe ser falsa: los continentes deben haberse movido." (p. 26) Esto le llevó a buscar la base empírica de lo que constituiría la inferencia abductiva de la hipótesis de la *deriva continental*: Datos Geodésicos, Datos Geofísicos, Datos Geológicos, Datos Paleontológicos y Datos Paleoclimáticos. Su conclusión final sería (1929, p.132. Mis cursivas, A. R.):

> *estas observaciones se han convertido en uno de los argumentos más convincentes a favor de la teoría movilista*.

6. La *preducción* en el contexto de descubrimiento en ciencias naturales teóricas

Las ciencias teóricas, en particular la física, se sirven de las matemáticas como herramienta de trabajo. Allí donde realizan sustituciones y combinaciones con magnitudes y ecuaciones, la legitimidad de las mismas la garantiza el análisis dimensional. Gracias a ello el físico puede *jugar* con fórmulas y símbolos con gran libertad y con el límite de que, donde los resultados de su juego son susceptibles de contrastación empírica, la Naturaleza tiene la última palabra sobre su utilidad o interés.

Como el físico teórico tiene siempre a su disposición toda la física, con todo su acervo de hipótesis, leyes y teorías, puede permitirse combinar las fórmulas aceptadas de las teorías, siempre que los dos miembros de la ecuación resultante sean dimensionalmente homogéneos. Esta práctica común en física teórica es lo que denomino *preducción*.

Veamos como ejemplo sencillo de preducción la propuesta de la naturaleza dual ondulatoria-corpuscular de los fotones, que Einstein obtuvo combinando resultados de teorías distintas: la relatividad especial y la física cuántica. Según la primera, la energía de los fotones viene dada por $E = c\,p$, donde *c* designa la velocidad de la luz en el vacío, y *p* el momento lineal o cantidad de movimiento de la partícula. Por otra parte, según Plank, la energía de un fotón es $E = h\nu$, donde *h* denota el *cuanto de acción*, y ν es la frecuencia del fotón. Pues bien, como ambas expresiones de la energía son dimensionalmente equivalentes, de su combinación matemática resulta que, $p = h/\lambda$ ó $\lambda = h/p$, e. d. las fórmulas que manifiestan el carácter dual de la radiación: el momento (propiedad corpuscular) de un fotón depende inversamente de su longitud de onda (propiedad ondulatoria); su longitud de onda depende inversamente de su momento.

La preducción teórica no es sino una implementación del método hipotético-deductivo. Su especificidad reside en que mientras el uso ordinario del método hipotético-deductivo va unido o bien al método axiomático-deductivo, o bien se le sitúa en el contexto de contrastación de teorías (contexto de justificación), o incluso en el contexto de explicación (Cfr. Rivadulla 2004, Cap. II), yo afirmo la extensión del método hipotético-deductivo también al contexto de descubrimiento científico. Es obvio pues que Peirce se equivocaba cuando afirmaba que el razonamiento deductivo no puede dar origen a nuevas ideas.

Entiendo pues como *preducción* la forma de razonamiento consistente en la combinación y manipulación matemática, de forma compatible con el análisis dimensional, de leyes del conjunto de la física, a fin de deducir resultados potencialmente interesantes. La preducción proporciona el procedimiento por el que muchos modelos teóricos, y muchas hipótesis de trabajo, son propuestos en ciencia, y por el que datos empíricos son también avanzados.

El razonamiento preductivo se diferencia de la abducción en que las hipótesis preducidas no vienen sugeridas por los datos, sino que se obtienen deductivamente a partir del acervo teórico disponible. Ahora bien, como se asume que éste puede no ser verdadero, la preducción constituye también una forma falible de manejarnos con la Naturaleza.

7. A modo de conclusión: El mito del método científico

Hemos visto en las páginas precedentes que las ciencias naturales observacionales se sirven, además de la inducción, del razonamiento abductivo para la propuesta de hipótesis explicativas. Y que, aunque éste no constituye tampoco una forma inferencial lógicamente válida, su uso se acepta como una *estrategia* de descubrimiento, por virtud de la cual proponemos hipótesis falibles, a fin de manejarnos científicamente con el Mundo. En efecto, los resultados de aplicar esta estrategia, es decir, las hipótesis abducidas, pueden ser erróneos. Peirce (CP, 2.777) mismo lo reconoce: "La hipótesis que [la abducción] problemáticamente concluye con frecuencia es totalmente falsa, y el procedimiento no necesita llevarnos siempre a la verdad".

Por lo que a la preducción respecta, las hipótesis, modelos y demás resultados preducidos, dependen de la totalidad del acervo teórico disponible, aunque no aceptado en cuanto verdadero. La preducción también hace posible la postulación de hipótesis falibles que nos permiten manejarnos predictivamente de modo falible con la Naturaleza.

Inducción, abducción y preducción, junto con analogía y serendipia, constituyen *estrategias de la creatividad científica para nuestro manejo falible con la Naturaleza.* Tal variedad de prácticas de descubrimiento científico falible ofrece el mejor apoyo a la tesis del mito del método.

Bibliografía

Carnap, Rudolf; Hans Hahn y Otto Neurath (1929). *Wissenschaftliche Weltauffassung. Der Wiener Kreis*. En H. Schleichert, *Logischer Empirismos-der Wiener Kreis.* Wilhelm Fink Verlag, München 1975.

Duhem, Pierre (1906). *La Théorie Physique. Son Objet et sa Structure*. Chevalier & Rivière, Paris.

Einstein, Albert (1936). "Física y realidad". En A. Einstein, *Mis ideas y opinions*. Bon Ton, Baercelona 2000.

Josephson, John R. (1994). "Conceptual analysis of abduction". In John R. Josephson and Susan G. Josephson (eds.), *Abductive inference. Computation, Philosophy, Technology*, Cambrige Univ. Press, New York 1994.

Mason, Ron (2001). "Stripes on the Sea Floor", en N. Oreskes (ed.), *Plate Tectonics,* West View Press.

Nagel, Ernest (1961). *The Structure of Science*. Routledge & Kegan Paul, London.

Peirce, Charles S. (1965). *Collected Papers*. Harvard University Press, Cambridge, Massachussetts.

Popper, Karl (1959). *The Logic of Scientific Discovery*. Hutchinson, London.

Popper, Karl (1966). *The Open Society and its Enemies*, Vol. II. Fifth ed. (revised). Routledge, London

Popper, Karl (1974). "Darwinism as a Metaphysical Research Programme", *Intellectual Autobiography.* En A. Schilpp (ed.), *The Philosophy of Karl Popper*

Putnam, H. (1981). *Reason, Truth and History*. Cambridge University Press

Reichenbach, Hans (1936). "Die Induktion als Methode derwissenschaftlichen Erkenntnis", *Actualités Scientifiques et Industrielles* 391

Reichenbach, Hans (1938). *Experience and Prediction*. Univ. of Chicago Press, Chicago

Rivadulla, Andrés (2003). *Revoluciones en Física*. Editorial Trotta, Madrid

Rivadulla, Andrés (2004). *Éxito, Razón y Cambio en Física. Un enfoque instrumental en teoría de la ciencia*. Editorial Trotta, Madrid

Rivadulla, Andrés (2006). "Metáforas y modelos en ciencia y filosofía". *Revista de Filosofía*, vol. 31, Num. 2, 189-203

Rivadulla, Andrés (2007). "Abductive Reasoning, Theoretical Preduction, and the Physical Way of Dealing Fallibly with Nature". En Olga Pombo & Alexander Gerner (eds), *Abduction and the Process of Scientific Discovery*, Centro de Filosofia das Ciências da Universidade de Lisboa.

Rivadulla, Andrés (2008). "Scientific Reasonableness and the Pragmatic Approach to the Unity of Science". En Olga Pombo, Shahid Rahman y Juan M. Torres (eds.), *The Unity of Science* (En prensa)

Roberts, Royston M. (2004). *Serendipia,* Alianza Editorial CT2514, Madrid

Russell, Bertrand (1949). *La perspectiva científica*. Ariel, Barcelona

Thagard, Paul R. (1978). "The Best Explanation", *Journal of Philosophy, vol.* 75, nº 2,

Wegener, Alfred L. (1929). *El origen de los continentes y océanos*. Ediciones Pirámide, Madrid 1983, de la segunda edición de 1929.

THE BEGINNING OF SCIENTIFIC CONTROVERSIES. SOME EPISTEMOLOGICAL AND LOGICAL CONSEQUENCES

Hassan Tahiri

Abstract. By using the latest results in nonmonotonic logic and in the history of science, this paper is aiming to make the three following points: 1. Historical: to challenge the received view concerning the history of astronomy according to which the Copernican revolution is the starting point of the scientific revolution. Recent researches in the history of astronomy, however, linked the Copernican system to the works of the 13 and 14th centuries Arabic eastern astronomers known as the Marāgha school whose aim was to find countermodels to the Ptolemaic system. 2. Epistemological: new historical facts show that the development of astronomy is due to controversies originated in the Arabic tradition indicating a major shift in scientific practice. In a previous paper, I showed that the elements on which the models of the Marāgha astronomers were build were provided by the work of the eleventh century Arabic physicist and mathematician Ibn al-Haytham. Furthermore Ibn al-Haytham'sepistemological work *al-Shukūk* (or *Doubts about Ptolemy*), in which he challenges the underlying assumptions of the *Almagest*, inaugurates a new way of doing science since it represents the first controversy in the history of science. The analysis of a *miniature*-example will illustrate the structure of Ibn al-Haytham exposition of the controversy, this follows the style of what later has been formalized as disputations or obligations with Ptolemy as a proponent and Ibn al-Haytham as an Opponent. More generally, scientific controversies can be seen as the driving force behind the development of science. 3. Logical: to introduce some formal concepts, based on the fundamental idea of nonmonotony, for the general treatment of scientific controversies. The aim is to develop a logical framework to be used as an instrument in understanding the development of scientific theories by refining the analysis of the different levels of argumentation.

The so called Copernican revolution is Kuhn's most cherished example in his conception of the non-cumulative development of science. Progress through Revolutions is the new paradigm by which Kuhn apparently succeeds in overthrowing the old one. Kuhn's famous revolution appears to be brought about by the gap between his

opponents' epistemological theory and the history of science since he presents his account of the story of *the Copernican Revolution* as if it is entirely based on facts. To what extent Kuhn's account is based on historical facts and Kuhn's epistemological theory captures the dynamic feature of the history of science? Can Kuhn's paradigm be overthrown? Can the gap, widened by Kuhn's epistemological theory, between philosophy and the history of science be bridged? And if so how? This paper is an attempt to answer some of these questions.

1. *The Copernican Revolution* revisited

A close examination of *The Copernican Revolution* reveals that Kuhn's reading of the history of astronomy is based on the following claims:

> **1.** The failure of Ptolemy's *Almagest* to reconcile theory and observation.
> **2.** The inevitable abandonment or replacement of the extant theory following its incompatibility with a single observation (p. 74) (and consequently his claim that the incompatibility of theory and observation is the ultimate source of every scientific revolution in the sciences (pp. 74-75).
> **3.** The perception of the imperfect nature of the *Almagest* by Ptolemy's successors prompts them to make several attempts to improve its accuracy.
> **4.** The attribution to Copernicus the invention of the mathematical theory needed for the development of the heliocentric idea.
> **5.** The revolution fails to materialise before because of extra scientific factors.

Kuhn goes even further. Indeed, in his view the Copernican model has not only introduced a major discontinuity in the history of science but the new paradigm and the old paradigm are incommensurable i.e. the gap between the two models are so huge that the changes introduced in the new model cannot be understood in terms of the concepts of the old one. By explicitly denying that the

development of the history of science could be internally understood, he denies that the history of science could have a dynamic of its own.

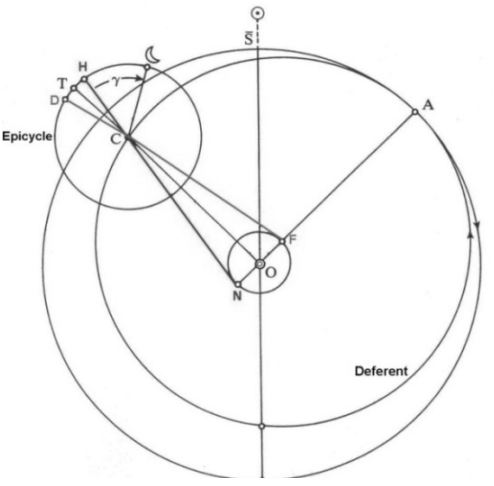
The full significance of the Arabic tradition appears in new light since it will be shown to be indispensable for our understanding of the development of science in general and of astronomy in particular. Indeed the study of Arabic science and philosophy can bridge the gap assumed by Kuhn as a historical fact precisely in the case of Copernicus. It will be shown that the changes involved in the work of Copernicus arouse as a result of an interweaving of epistemological and mathematical controversies originated in the Arabic tradition which challenged the Ptolemaic model nine centuries later. My main case study is the work of the eleventh century Arabic scientist Ibn al-Haytham who devoted a whole book to the task of refuting the implications of the *Almagest* machinery, a landmark attempt in the history of science since its resounding and overwhelming success was decisive for the progress of science. Ibn al-Haytham's *al-Shukūk* had such an impact that since its disclosure the *Almagest* stopped to be seen as the suitable model of the heavenly bodies. Numerous attempts have been made to find new alternative models based on the correct principles of physics following the strong appeal by both Ibn al-Haytham and, after him, Ibn Rushd. The work of the astronomer Ibn al-Shātir, based exclusively on the concept of uniform circular motion, represents the climax of the intense theoretical research undertaken during the thirteenth and the fourteenth centuries by Arabic and Islamic astronomers of widely known as Marāgha School (which owes its name to the observatory of Marāgha in north-western Iran). I argue that the point of the connection between the works of Ibn al-Haytham and Ibn al-Shātir is that while the *al-Shukūk* gives the elements to build a countermodel to the *Almagest,* the work of Ibn al-Shātir offers a model which, three centuries later, takes care of the objections triggered by the work of Ibn al-Haytham. Furthermore, not only the basic identity of the models of Ibn al-Shātir and Copernicus has been established by recent research, but that it was also found that Copernicus used the very same mathematical apparatus which were developed by the Marāgha School during at least two centuries. It turns out that Copernicus was in fact working under the influence of the two streams of the Arabic tradition: the well known more philosophical western stream, known as physical realism and the newly discovered eastern mathematical stream. The first relates to the

idea that astronomy must be based on the principles of physics and physics is about the real nature of things. The second relates to the use of mathematics in the construction of models and countermodels in astronomy as developed by the Marāgha School. This paper will focus mainly on the recently discovered eastern mathematical stream since the western philosophical influence is well known and can be found in Duhem' s monumental work *Le Système du Monde*. The case presented is a further challenge to the role of a mere intermediary status assigned to the Arabic tradition by the standard interpretation of the history of science and more generally it presents a first step towards a reconsideration of the thesis of the discontinuity in the history of science. My view is that major changes in the development of science might sometimes be non-cumulative though this is not a case against continuity understood as the result of a constant interweaving of net of controversies inside and beyond the science at stake.

2. The beginning of scientific controversies over the foundations of astronomy: Ibn al-Haytham's *Doubts against Ptolemy's Almagest*

The object of the controversy concerns the motion of the planets which is by far the main controversial issue raised by *al-Shukūk*.

2.1. Ptolemy's account for the motion of the moon: the introduction of the prosneusis hypothesis



$\bar{S}$ is the mean apogee (i.e. a fictitious body that moves uniformly around the earth O); F is the centre of the deferent; C is the centre of the epicycle on which the moon moves uniformly; T, the apogee seen from O, is called the true apogee; H, the point from which the mean anomaly is measured, is called the mean apogee; N is the prosneusis point.

This description of the motion of the moon consists of the following features:

> **(i)** the deferent moves around its centre F such that $\bar{S}$ OA and $\bar{S}$ OC are equal and opposite, we have here a circular body moving uniformly around a point O other than its centre (this is the main feature of the equant hypothesis);
> **(ii)** instead of counting the anomaly γ which determines the distance of the moon M on the epicycle from the apogee D, it had to be measured from the point H (called for this reason the mean apogee) such that the radius HC has a direction towards a point N which is always located diametrically opposite to the point F;
> **(iii)** H is a variable apogee since N, the prosneusis point, has to be always in motion to remain opposite to the moving point F. The difference thus between the equant and the prosneusis is that the description of the motion of the moon starts from a non-stable point. We call the prosneusis hypothesis the assumption of moving the line HN instead of FD;

One of the major consequences of Ptolemy's work is that it puts an end to the unclear relationship between mathematics and physics. His decision is to subordinate the latter to the former; this philosophical approach involves a particular conception of science which has, as one of its severe side effects, widened the gap between mathematics and physics. It remains to be seen whether it is the most fruitful approach. These difficult questions are not raised by the author of the *Almagest*. Given Ptolemy's exposition of his theory, the reader should not expect from the mathematician-astronomer to discuss its philosophical and epistemological implications. A controversy is

needed to challenge its underlying assumptions and to bring to the forefront these foundational issues.

Let us follow in a *miniature*-example the structure of Ibn al-Haytham exposition of the controversy. By the way, this follows the style of what later has been formalized as disputations or obligations with Ptolemy as a proponent and Ibn al-Haytham as an Opponent.

2.2. *The status of the prosneusis hypothesis*

We open the dialogue by Ptolemy's main thesis: in his first argument Ptolemy justifies the prosneusis hypothesis required by the motion of the lunar diameter (and the whole the *Almagest* machinery) by the aim of accounting for the phenomena.

Ibn al-Haytham begins his counterargument by noting that according to the uniform circular motion principle, a circular body should move uniformly around its centre. The assumption of the motion of HN (i.e. the prosneusis hypothesis) means that the centre of the epicycle does not move uniformly around the deferent. Therefore, the prosneusis hypothesis contradicts the uniform circular motion.

Ptolemy seems to have anticipated this objection by adding the following argument: there is no contradiction between the prosneusis hypothesis and the uniform circular motion principle since HN is an imaginary line.

Ibn al-Haytham replies swiftly and surprisingly: but an imaginary line does not move; an argument which reinstates his objection. For Ibn al-Haytham to assume thus the existence of imaginary objects in motion is an absurdity.

Can the motion of the lunar diameter be justified? And if so how?

2.3. The justification of the prosneusis hypothesis

Ibn al-Haytham points out that Ptolemy has considered an assumption in his *Planetary Hypotheses* in which he introduces a body (a sphere or a disc) that moves the epicycle. Then he shows in a long argument that the prosneusis hypothesis needs itself another assumption: an entity which can perform not only two kinds of

motion, it is also required from the two motions to be carried out in opposite directions. It is in this context that we have to understand why the most virulent attack ever launched by Ibn al-Haytham against Ptolemy's arguments is the one involving of what can be called the second order assumption.

And the *inquirers among philosophers believe* من الفلاسفة يعتقدون و المحقّقون that no two opposite motions can exist in the heavens. It is one of the most absurd impossibilities فمن أفحش المحال that they can be possessed by one and the same body of the heavenly bodies (My emphasis, p. 37).

This is not the only place where Ibn al-Haytham mentions explicitly the philosophers. He also appeals to them when he discusses a similar point. The aim of Ibn al-Haytham' s argument is to establish the proposition that each body must have only one kind of motion is a second fundamental principle of motion (F.P.M.); the first being uniform circular motion. It is now easy for him to show by reductio ad absurdum that the assumed body required by the prosneusis hypothesis cannot exist.

(i) Assume that there is such a body B;
(ii) B must have two opposite motions;
(iii) But a body can have only one kind of motion;
(iv) Therefore B cannot exist.

And from the non-existence of B follows that the prosneusis hypothesis cannot be justified.

By bringing in the philosophers into the controversy, Ibn al-Haytham no doubt hopes they will take up the matter more deeply; this is a clever move aimed at radicalising further the controversy, if it succeeds, can only speed up the collapse of Ptolemy' s system by bringing to the surface all its hidden assumptions and weaknesses.

Al-Shukūk is not simply a book about astronomy as it is generally believed but it is much more profound than that. It is an unprecedented philosophical and epistemological doctrine on how progress in science can be achieved. The book is clearly divided into two parts: an extremely condensed and poetical (to ensure that his discourse will have as much effect as possible) introduction in which Ibn al-Haytham exposes his dynamic approach to the development of science. The rest of the book is an implementation of this approach to

astronomy, the most advanced scientific discipline. *Al-Shukūk* was widely known and quoted both in eastern and western Arabic countries. His strong counterarguments to the *Almagest*'s hypotheses have the desire effect since his forceful appeal has been answered by both astronomers and philosophers.

3. The impact of *Al-Shukūk*: the deepening of controversies over the foundations of astronomy

The Arabic western philosophers have deepened the controversy on the foundations of astronomy by explicitly rejecting out of hand the *Almagest* machinery. To illustrate such an outright rejection by the western tradition, we present the final verdict of Ibn Rushd taken from his *Commentary on Aristotle's Metaphysics* in which he condemns once and for all the Ptolemaic system.

> The astronomer should build an astronomical system from which the heavenly motions follow and such that there is no impossibility from the physical standpoint … Ptolemy has not succeeded in putting astronomy on its true foundations. The epicycle and the eccentric are impossible. It is then necessary to conduct new research for this true astronomy whose foundations are the principles of physics. In my youth I had hoped to accomplish this investigation, but now in my old age I have despaired of that, having been impeded by obstacles.

And he ends up with the same conclusion as that of Ibn al-Haytham

> But let this discourse spur someone else to inquire into these matters [further]. For nothing of the true science of astronomy exists in our time, the astronomy of our time being only in agreement with calculations and not with what exists. (in Duhem 1913, volumeII p. 138)

The new astronomy should thus be founded on the correct principles of physics: this task is left to younger generation of astronomers. It is in this context that al-Bitrūjī composes his *on the*

Principles of Astronomy in which he proposes a new conception of the universe based on the principles of dynamics.

This is the well known influence of the western Arabic philosophical tradition. Recent research has demonstrated that Copernicus was also working under the influence of the eastern Arabic mathematical tradition. The discovery from 1957 onwards of the writings of a series of astronomers of the Marāgha School of the thirteenth century shows that the fever of Ibn al-Haytham's *Al-Shukūk* has also spread further to the East. The work of the Marāgha astronomers was motivated by the same aim: to find counter-models to the Ptolemaic system. Their results have far reaching consequences on our understanding of the history of astronomy. George Saliba, one of the scholars who closely studies these works, describes here the sensation caused by the discovery

> First, the Marāgha results were, from the very beginning, deemed extremely important on account of their relationship to the works of Copernicus. But one should say that this relationship did not touch upon the Copernican notion of heliocentricity. That feature of Copernican astronomy entails the transformation of geocentric mathematical models to heliocentric ones by the reversal of the vector connecting the sun to the earth, while leaving the rest of the mathematical models intact. It is rather the similarity of the Copernican geocentric versions of those models to those of the Marāgha astronomers that invited curiosity. (idib. p. 265-267)

The sensation in these new findings is not so much the identity of the Copernicus models and those developed by the Marāgha astronomers three centuries earlier since it is perfectly conceivable that Copernicus could have developed independently his own models. Striking is the fact that Copernicus uses without proof some mathematical results already proven (geometrically) by the Marāgha School three centuries before. Saliba explains what is at stake on the historical level as far as the development of astronomy is concerned.

> What is clear is that the equivalent model of Ibn al-Shātir seems to have a well established history within the results reached by earlier Muslim astronomers, and could therefore be historically explained as a natural and gradual development that had started

> some three centuries earlier. The same could not be said of the Copernican model (Saliba 1994, p. 304; also Saliba 1996, p. 113).

And he concludes that the Copernican astronomy cannot be very well understood, on the mathematical and technical levels, without careful study of the achievements of the earlier Marāgha astronomers. The time is now ready for the historians of civilisations to step in. The eminent historian of ancient as well as medieval astronomy Neugebauer has this to say

> The recovery of the planetary theory of the astronomers of the Marāgha School […] is not only of great interest in itself, but has also demonstrated that much what had been taken for Copernicus's own planetary theory is actually of medieval Arabic origin, and was transmitted to western Europe by an unknown route, perhaps by way of late Byzantine sources, to Italy at some time in the fifteenth century. (Neugebauer 1984, p. 290)

Neugebauer adds "the question therefore is not whether but when, where, and in what form, he [Copernicus] learned of Marāgha theory" (ibid., p. 47). It looks like as if the history of science is about to be rewritten. The status of Copernicus seems to be more hanging in the balance than ever, blurring thus the sharp distinction to which he is traditionally associated with. The received view finds in the *De Revolutionibus* a clear-cut between medieval European era and the renaissance since it contains some significant results which make his author the starting point of a revolution. But on the face of this new evidence, it seems that the received view on the periodisation of science, already challenged by Duhem, needs to be revised as it is suggested by the following statement of Neugebauer: "in a very real sense, Copernicus can be looked upon, as if not the last, surely the most noted follower of the Maragha School" (Neugebauer 1984, p. 47).

What is remarkable is that the eleventh century Arabic philosopher and scientist has successfully answered the current epistemological question concerning the progress of science. His philosophical doctrine is that a scientific theory reaches a point where

controversies are necessary for its further development. Contrary to his *Optics*, *al-Shukūk* is not a book of science but a book about science since it is motivated most and foremost by epistemological considerations designed to break the deadlock caused by the Ptolemaic exposition of science. By systematically challenging the various works of Ptolemy, Ibn al-Haytham demonstrates how progress in science can be achieved since his *Optics* represents the first successful attempt in overthrowing an existing scientific theory and his *al-Shukūk,* which marks a turning point in the history of astronomy, represents the first successful attempt in demonstrating the power of argumentative analysis in bringing about a fundamental shift in the development of a scientific theory. This case study not only calls for a serious reconsideration of the thesis of discontinuity but more generally should prompt historians of science and philosophy to revise the prevailing periodisation of science. Furthermore, the refutation of Kuhn's theory paves the way for bridging the gap between philosophy and the history of science. More precisely, the new and intrinsic dynamic interpretation of the history of science can be captured by the analysis of scientific controversies by using the logic of argumentation which is based on the idea of nonmonotony.

4. Basic concepts of the logic of argumentation

Nonmonotonic reasoning is a reasoning in which temporary conclusions can be drawn on the basis of incomplete information but which might be withdrawn when more information becomes available. Systems for defeasible argumentation capture this kind of reasoning in terms of interactions between conflictual arguments. Nonmonotonic reasoning is explained in terms of defeasibility since arguments can be defeated by stronger counterarguments.

The logic of argumentation is set up on three fundamental concepts: the notion of argument, the conflict between arguments and the resolution of the conflict.

4.1. The notion of argument and its structure

The purpose of the logic of argumentation is not to define what is an argument. That is why systems for defeasible argumentation

need a logical language which specifies the formation rules of arguments. The notion of logical consequence of this underlying language is monotonic in the sense that new premises cannot invalidate an argument as such but can only give rise to a counterargument. Nonmonotonicity arises when the added new premises give rise to a stronger counterargument which defeats the original one.

An argument is composed of two parts:

The first part consists of a set of premises or subarguments;
The second part is its conclusion

The subarguments provide reasons which justify the conclusion. There are two kind of rules from which subarguments can be constructed:

Strict or monotonic rules:

$$A \rightarrow B$$

Defeasible or nonmonotonic rules:

$$A \Rightarrow B$$

An assumption (or defeasible information) can be used in the derivation of a conclusion, this is translated symbolically by a defeasible rule with zero subargument $\Rightarrow A$.

These two types of rule give rise to the formation of two kinds of argument:

An argument is defeasible if one of its subarguments is defeasible; otherwise, it is a strict argument.

It should be noted that in nonmonotony, it is the conclusion of an argument which is the object of the dispute, i.e. the inference of a proposition from a given number of premises. And the task of the logic of argumentation is the study of the relation between conflictual arguments.

4.2. The conflict between arguments

As mentioned above, an argument stays an argument if more premises are added. If adding new premises cannot invalidate an argument, the latter can however be disputed by another argument. It is thus important to know when arguments are in conflict. We will present briefly the ways in which two arguments can be in conflict with each other.

4.2.1. Rebutting an argument.

This type of conflict occurs when arguments have contradictory conclusions as in the following example.

Proposant: $A \Rightarrow B$
Opposant: $C \Rightarrow \neg B$

The proponent asserts if A then B, the opponent rebuts the proponent's conclusion by appealing to an argument C. Here is an intuitive example taken from biology where new species are constantly discovered

Proponent	X flies because it is a bird
Opponent	X does not fly because it is a penguin

4.2.2. Assumption attack

Two cases should be distinguished here. The first case in which the assumption is explicit. In this case, we have an assumption attack when one argument makes a non-provability assumption (as in default logic) and another argument proves what was assumed unprovable by the first argument. This is the formalisation of this kind of attacks:

P: $A \wedge \sim C \Rightarrow B$
O: $D \Rightarrow C$

The availability of new information (D) may provide evidence to the contrary of a previously acceptable default conclusion which has now to be withdrawn. By defeating the assumption of non-provability ~C, the opponent prevents the inference of the proposition B.

For example, "X flies because it is a bird and it is not provable that X is a penguin" is attacked by an argument the conclusion of which is "X is a penguin".

To show you how far the logic of argumentation can really capture common sense reasoning, I am tempted here to give you a live example which should be recorded in historical books. In an effort to justify his war on Iraq, President Bush has all along argued:

Iraq has weapons of mass destruction because it used them in the past.

Bush's main claim assumes the following argument:

There is no evidence that the weapons of mass destruction were destroyed.

But after years of research conducted by the inspectors of the UN, Saddam was able to attack with much strength Bush' s non-provability assumption:

The weapons of mass destruction were destroyed in the early nineteen nineties.

The controversy did not stop there since Saddam's argument did not prevent the invasion of Iraq. Why? Just because Saddam'sargument is not a strict argument: his opponent required from him to provide evidence of the destruction of weapons of mass destruction. This explains why Bush managed to find more arguments designed to weaken his opponent'sstrong position; let me give you just one of the many counteraguments used by Bush namely the one that cast doubt on the credibility of his opponent. But what is more interesting as far we are concerned is once Bush was forced to swallow his sworn enemy'sargument, his main claim disappeared mysteriously from his subsequent speeches admitting in effect that he

lost the argument. This outcome had very serious implications since it prompted the collapse of Bush'soverall strategy leaving his country, Iraq and unfortunately the rest of the world in a very critical situation.

The attack on an explicit assumption has the same effect in the second case where the assumption is implicit. In the following move, the proponent asserts that B is justified by A.

P: $A \Rightarrow B$

The aim of the opponent here is to attack neither A nor B but an implicit assumption C used in inferring B. In this case, he should make explicit the argument C by reformulating the full argument of the proponent designated here by P' before proceeding to attack the implicit assumption.

P': $A \wedge C \Rightarrow B$
O: $D \Rightarrow \neg C$

4.2.3. Undercutting an inference

What is common between the first two kinds of conflict is that arguments attack statements. The third type of conflict is an attack not on a statement but on a rule of inference. Only non-deductive rules can be subject to an inference attack. Argumentation systems can thus allow the use of non-deductive reasoning such as inductive, abductive or analogical arguments.

This is an example of a non-deductive argument: "$raven_{101}$ is black since the observed ravens: $raven_1$ … $raven_{100}$ were black", this inductive reasoning is undercut by the argument "I saw a raven which was white". So may be $raven_{101}$ could not be black after all. Here is a concrete example of undercutting an inference which has been successfully used in the Netherlands. Euthanasia is a punishable act because it is considered as a murder. But doctors can be faced with a situation in which on the one hand they have the duty to preserve life and on the other the duty to respect a person'sexplicit wish to die. In the Netherlands, such a dilemma is considered as a force majeure which constitutes a legal ground for not prosecuting. So the general rule which says that "if someone helps a person to commit suicide then he must be punished" can be undercut by the exception rule "if a

doctor helps a patient to commit suicide and the patient explicitly expresses his wish to die then he acts in force majeure".

The three types of attack can be used in the same controversy. It is the nature of argumentation which determines both the choice and the mode of expressing arguments and counterarguments.

4.3. The resolution of the conflict between arguments

To resolve the conflict between two arguments, the approach of nonmonotonic logic is to evaluate their relative strength in respect with each other. Three methods are used in this regard.

Priority rules

The first priority used in nonmonotonic logic for a long time is the specificity rule which proved to be very useful in empirical sciences. Here the idea is that more an argument is specific more it is informative and thus reliable.

Proponent: X flies because it is a bird
Opponent: X does not fly because it is a penguin

O defeats P in virtue of the specificity rule.

Priority rules are heavily used in law to resolve possible conflictual laws. Among many priority rules a priori adopted, let us mention the rule which stipulates that in case of a conflict between two or several laws, it is the most recent law which should be applied.

Not only arguments, but also the type of attacks used by the parties themselves can enter in a conflict as the following example shows

$$P: \sim \neg A \Rightarrow A$$
$$O: \quad \Rightarrow \neg A$$

Here it seems that we have a rebuttal attack since the two arguments have contradictory conclusions. But looking more closely to the structure of both arguments, what the opponent actually attacks is the premise rather than the conclusion of the proponent'sargument.

P asserts that he can infer A by assuming that there is no proof of ¬A. But O replies now that he can prove ¬A; by attacking the non-provability assumption used by the proponent, the opponent prevents the inference of A. This explains why in case of conflict between a rebuttal and an assumption attack, it is the latter which has the upper hand. In other words, attacking a subargument is stronger than rebutting an argument which consists merely in attacking its conclusion.

A third possibility is that the priority rules are not fixed in advance but are themselves subjected to arguments and counterarguments between the parties. The underlying idea here is that the criteria which should serve as an argument of priority are determined by the subject domain.

The burden of proof

In classical dialogues such as e.g. monotonic dialogues, the proponent has always the burden of proof since he wants to justify his claim. The opponent role is conceived as a guide for justifying the proponent'sclaim or at best merely to try to contradict him (giving possibly rise to a countermodel). The burden of proof is static here since it cannot change from one side to another. But in defeasible argumentation, the burden of proof can change sides since the opponent can add new premises, make claims, assumptions, … which can be attacked by the proponent. To capture the dynamic feature of the burden of proof that characterises procedural systems, a double distinction in the role played by the parties should be introduced. The first distinction concerns the classical role of the players in the dialogue which can be called the dialectical role: proponent/opponent. The dialectical role remains fix throughout the whole dialogue since it is determined with respect to the main claim. This role should be distinguished from the procedural role of the parties which is determined with respect to the burden of proof: attacker/defendant. Except at the start of the dialogue, the burden of proof can be attributed either to the proponent or the opponent depending on the development of the exchange of arguments.

The audience

The third criterion that can be used to break the deadlock between the parties concerning the acceptation of a given argument is to refer to the scientific community. Indeed the audience is the real raison d' être (since it provides the background from which controversies are triggered) and at the same time the target of controversies. The ultimate aim of each player is to convince his audience since the party who triumphs is usually measured by the fact that his thesis succeeds in gaining the support of the majority of the members of the scientific community. But the powerful nature of controversies, which explains the fact they represent the driving force behind the development of science, lies not only in succeeding to have an immediate influence on the current audience but also in their ability to sustain its huge impact far beyond the spatial-temporal context that gives them birth in the first place. This is the lesson to be learned from our present case study.

References

• Primary sources

Ibn al-Haytham (1971). *al-Shukūk 'ala Batlamyus*, Edition of the Arabic text by A.I. Sabra and N. Shehaby, Dar al-Kutub, Cairo.

Ptolemy, C. (1984). *The Almagest*, Translated and annotated by G. J. Toomer, Duckworth, Classical, Medieval and Renaissance editions, London.

• Secondary sources

Al-Bayhaqi (1946). *Tarikh Hukama' al-Islam*, al-Taraqqi Press, Damascus.

Al-Bitruji (1971). *On the Principles of Astronomy*, Translated and annotated by B. Goldstein, 2 volumes, New Haven, Yale University Press.

Carra, de V. (1893). ' Les sphères célestes selon Nasīr Eddīn-Attūsī' , in P. Tannery, *Recherches sur l' histoire de l' astronomie ancienne*, appendice VI, pp. 337-361, Editions Jacques Gabay.

Dreyer, J. L. M. (1953). *History of Astronomy from Thales to Kepler*, Dover. (a replication of the 1906 *History the Planetary Systems from Thales to Kepler*, Cambridge University Press.)

Duhem, P. (1913-1965). *Le Système du monde: Histoire des doctrines cosmologiques de Platon à Copernic*, 10 volumes, Hermann, Paris.

Duhem, P. (1994). *ΣΩZEIN TA ΦAINOMENA*: *Essai sur la notion de théorie physique*, Mathesis, Vrin, Paris.

Evans, J. (1998). *The History and Practice of Ancient Astronomy*, Oxford University Press.

Hebraeus, B. (1899). *Le livre de l ascension de l' esprit*, Traduction de F. Nau, Bouillon, Paris.

Keostler, A. (1968). *The Sleepwalkers,* The Danube Edition, Hutchinson of London.

Kuhn, T. S. (1957). *The Copernican Revolution: Planetary Astronomy in the Development of Western Thought*, Harvard University Press.

Lettinck, P. (1994). *Aristotle's Physics and its reception in the Arabic World with an Unpublished parts of Ibn Bājja's Commentary on the Physics*, Leiden, New York, Köln.

Lettinck, P. (1999). *Aristotle's Meteorology and Its Reception in the Arab World: With an Edition and Translation of Ibn Suwar's Treatise on Meteorological Phenomena and Ibn Bājja's Commentary on the Meteorology,* Brill Academic Publishers.

Morélon, R. (2000). ' Ibn al-Haytham et ses arguments cosmologiques' in *Cosmologie et Philosophie* [Texte imprimé] = *Cosmology and philosophy*: hommage à Jacques Merleau-Ponty: [actes] du colloque / réalisé avec le concours du Centre de Philosophie et d' histoire des sciences de l'Université de Paris 10-Nanterre, l' Équipe REHSEIS (UMR, CNRS et l'Université Paris 7 Denis Diderot) et le Centre d' Histoire des sciences cosmologiques de l' Université de Reims, Champagne-Ardennes, Paris-Nanterre, 6-8 octobre 1997; sous la direction de J. Seidengart & J.-J. Szczeciniarz ; Epistémologiques, pp. 101-111.

Neugebauer, O., and Swerdlow, N. M. (1984). *Mathematical Astronomy in Copernicus's De Revoluionibus*, 2 volumes, Springer-Verlag.

Neugebauer, O. (1957). *The Exact Sciences in Antiquity*, second edition, Brown University Press.

Pines, S. (1964). 'Ibn al-Haytham'scritique of Ptolemy', *Actes du dixième congrès international des sciences* I **10**, pp. 547-550, Paris.

Rahman, S. Gabbay, D., Symons, J. and J.P. van Bendegem (Ed.) (2004). *Logic, Epistemology and the Unity of Science*. Dordrecht: Springer (Kluwer).

Rahman, S. and Symons, J. 'Logic, Epistemology and the Unity of Science: An Enyclopaedic Project in the Spirit of Neurath and Diderot' in Rahman, S., Gabbay, D., Symons, J. and J.P. van Bendegem (Ed.) 2004, pp. 3-17.

Rahman S., T. Street, H. Tahiri (2008). *The Unity of Science in the Arabic Tradition: Science, Logic, Epistemology and their connections*, Springer.

R. Rashed and R. Morélon (1996). *an Encyclopedia of the History of Arabic Science*, 3 volumes, Routledge.

Sabra, A. I. (1994). *Optics, Astronomy and Logic, Studies in Arabic Sciences and Philosophy*, Aldershot-Variorum.

_____. (1998). 'Configurating the universe: aporetic, problem solving and kinematic', *Perspectives on Science* **6**, pp. 288-330.

Tahiri, H. (2008). The Birth of Scientific Controversies, The Dynamics of the Arabic Tradition and its Impact on the Development of Science: Ibn al-Haytham'sChallenge of Ptolemy's*Almagest* in S. Rahman, T. Street, H. Tahiri: *The Unity of Science in the Arabic Tradition: Science, Logic, Epistemology and their Connections*, pp. 183-225, Springer.

(2005). *La dynamique de la négation et la logique avec inconsistances : quelques conséquences scientifiques et épistémologiques, vers un rapprochement entre la philosophie et l' histoire des sciences,* Université de Lille 3 (Doctoral Thesis).

"AND THEN THERE WERE NONE" Y EL RAZONAMIENTO ABDUCTIVO. EL PODER ESCLARECEDOR DE LA LÓGICA

Corina Yoris-Villasana
Universidad Católica Andrés Bello, Caracas

Resumen. Acercarse a los razonamientos abductivos con la idea de aplicarlos a un estudio literario es, por lo menos, audaz. Es un lugar común entre los literatos decir que la lógica no tiene nada que ver con la "buena literatura". Para muchos de ellos, el frío escalpelo lógico no posee cabida en la esfera del análisis de una determinada obra. Sin embargo, acerquémonos a una novela policial y veremos cómo el éxito de ella va a depender de la manera cómo se maneje el recurso racional para llegar a la solución del enigma planteado. ¿Cuáles son los elementos básicos de una novela policial? En primer lugar, hay un delito, sea éste un crimen, un asalto, que plantea de inmediato un ¿quién? que llevó a cabo el "qué"; por supuesto hay un móvil, sean éstos codicia, ira, celos, que responderían a la pregunta ¿por qué?, y, por supuesto, los medios mediante los cuáles se cometió el acto criminal. Éstos serán el ¿cómo? que esbozan, de una u otra forma, el momento, el contexto, es decir, el ¿cuándo?
Veamos, por ejemplo, **AND THEN THERE WERE NONE** de Agatha Christie. Nos centraremos en el razonamiento del detective para esclarecer el crimen. Hay un qué: el asesinato de los diez invitados; un cómo: las diversas maneras de la suerte de estos invitados. Poco a poco, se llega al porqué y el ¿cuándo? Todos estos elementos nos conducen al ¿quién?, que, en definitiva, es el propósito básico de la novela. Es el clásico **Who-do-it** inglés.
¿Qué tipo de razonamiento es el usado? Ése es nuestro objetivo. Mostrar cómo se llega al desenlace mediante el uso de operaciones lógicas que, a su vez, nos llevan a hipótesis "novedosas", las cuales permiten encontrar la solución.

Abstract. To try and understand abductive arguments with the idea of applying them to literature is, to say the least, audacious. It is commonplace among those who write literature as a profession that logic has nothing to do with literature. For many, logic has no business when analysing a work of fiction. Nonetheless, lets analyse the police novel and we will see how the success of it will depend on how the writer uses the rational recourse to arrive to the solution of the problem. What are the basic elements of a police

novel? First, there is a crime, then there is a "who?" who did the "what" then there is a motive and the means used to commit the crime.
We will study Agatha Christie's **And Then There Were None**. We will center our analysis in the detective'sreasoning to solve the crime. There is a what: the killing of ten guests; there is a how: the different ways in which they were killed; and then there is a why and a when. All of these elements will take us to "who?" which is the central problem of the novel. It is a classical "whodunit
What kind of reasoning is used? This is the main purpose of this essay: To show how we arrive to the end through the use of logical operations which take us to a hypothesis which in turn will let us arrive to a conclusion.

Keywords: Abductive reasoning, Police novel, Ten Little Niggers

Introducción

En cualquier manual escrito sobre novelas policiales, podemos encontrar la descripción elemental del género; básicamente consiste en un relato donde hay un hecho criminal que presenta un enigma y éste debe ser resuelto. Hay una suerte de confrontación entre quien narra y quien lee; éste tiene un reto por delante: encontrar una respuesta creíble al problema planteado. Para ello, el discurso que se emplee debe responder a una determinada estructura absolutamente lógica. En este trabajo nos proponemos efectuar un análisis, precisamente desde la perspectiva lógica, del razonamiento realizado por los policías de Scotland Yard, personajes de una las novelas más vendidas de la escritora británica Agatha Christie, escrita en 1939 bajo el título *Ten Little Niggers*, conocida también como *And Then There Were None*[1] para ¿resolver? el acertijo ideado magistralmente por la autora.

La trama es la siguiente: diez personas, que ocultan hechos delictivos no castigados por la justicia, son invitadas a una isla solitaria por un desconocido que se hace llamar a sí mismo Mr. U. N. Owen (adviértase que es una mezcla de letras del vocablo Unknown) y serán asesinados uno a uno llevando a cabo las pautas sugeridas en una canción de cuna[2]. Es de capital importancia leer atentamente la

[1] CHRISTIE, Agatha. *Diez Negritos.* Barcelona, Editorial Molino, 2004. Traducción de Orestes Llorens. Las citas corresponden a esta edición en español.
[2] Diez negritos se fueron a cenar. Uno de ellos se asfixió y quedaron Nueve// Nueve negritos trasnocharon mucho. Uno de ellos no se pudo despertar y quedaron Ocho// Ocho negritos viajaron por el Devon. Uno de ellos se escapó y quedaron Siete// Siete

letra de la canción porque allí está la clave para resolver el enigma. Los personajes son:

1. Anthony Marston, conduciendo de manera imprudente arrolló a John y a Lucy Combes. Debido a su riqueza y posición social, nunca fue juzgado y tan sólo le fue retirada su licencia de conducir.
2. Mr. and Mrs. Rogers dejaron morir a su patrona inválida, Jennifer Brady, ocultándole su medicina.
3. General MacArthur envió al amante de su esposa a una misión suicida durante la Primera Guerra Mundial.
4. Miss Brent despidió a su criada, Beatrice Taylor, cuando supo que estaba embarazada. La criada se suicidó.
5. Wargrave condenó a muerte a Edward Seton, a pesar de la evidencia de su inocencia.
6. Armstrong operó embriagado a Louisa Mary Clees ocasionándole la muerte.
7. Blore plantó falsa evidencia en el juicio por el robo a un banco a James Landor, quien más tarde murió en prisión.
8. Lombard abandonó a un grupo de 21 nativos ocasionándoles la muerte en un bosque del África.
9. Vera Claythorne voluntariamente permitió que Cyril Hamilton, un pequeño que estaba a su cuidado, nadara mar afuera y se ahogó.

Una de las características de Agatha Christie es su honestidad como escritora: ella nos está dando la clave para solucionar el problema. La novela culmina con dos capítulos absolutamente asombrosos. Ellos por sí solos constituyen otra novela policial. El epílogo muestra cómo los investigadores son incapaces de descifrar qué ha acontecido en la isla; no saben quién ha matado a quién; cuál fue el orden en que se sucedieron tales muertes, los motivos que tenía

---

negritos cortaron leña con un hacha. Uno se cortó en dos y quedaron Seis// Seis negritos jugaron con una avispa. A uno de ellos le picó y quedaron Cinco// Cinco negritos estudiaron derecho. Uno de ellos se doctoró y quedaron Cuatro// Cuatro negritos fueron a nadar. Uno de ellos se ahogó y quedaron Tres// Tres negritos se pasearon por el Zoológico. Un oso les atacó y quedaron Dos// Dos negritos se sentaron a tomar el sol. Uno de ellos se quemó y quedó nada más que Uno// Un negrito se encontraba solo. Y se ahorcó y no quedó... ¡Ninguno!

el asesino, y mucho menos quién es éste. Pero…en el último capítulo, titulado *Documento Manuscrito enviado a Scotland Yard por el Capitán del Pesquero Emma Jane,* el asesino, quien no resiste la tentación de alardear sobre su ingenio, envía un mensaje en una botella donde se aclara todo el caso.

Son estos dos capítulos donde centraremos nuestro estudio, pues es allí donde los argumentos empleados son los que nos sugirieron realizar el análisis propuesto en líneas precedentes. Generalmente, la tradicional historia de detectives termina con una escena en la que el detective, después de examinar exhaustivamente todas las pruebas, reúne al resto de personajes y explica todo lo que ha sucedido, de tal manera que concluye desenmascarando al asesino. Sin embargo, en este caso no es así. Los detectives de la policía quedan totalmente desconcertados por lo que ha sucedido, y será otro personaje quien explique las cosas y desenmarañe el misterio.

Delimitación del problema

> Sir Thomas Legge, subjefe de policía de Scotland Yard, decía enfadado:
> - Pero, ¡esa historia es increíble!
> - Ya lo sé, señor – respondió deferente el inspector Maine.
> - ¡Diez personas muertas y ningún ser viviente en una isla! ¡Eso es absurdo!
> - Sin embargo*, ha ocurrido*, señor – respondió impasible el inspector Maine.
> - ¡Maldita sea, Maine, alguien debe haberlas matado!
> - Eso es precisamente **nuestro problema**, señor.
> (A. Christie, 2004, 171. Negritas nuestras)

Así comienza el *Epílogo* de la novela. Los dos detectives, llamados para descifrar el crimen, se encuentran deliberando sobre el caso. No hay alusión alguna al sitio donde se encuentran. Podemos suponer que están en alguna oficina en el pueblo cercano a la isla.

Hay un problema para resolver: diez muertos en una isla desierta y ninguna persona en el lugar. Alguien las mató; es decir, tenemos el **qué**; hay que buscar el **quién**.

Razonamiento:
Premisa: Hay diez personas muertas en una isla desierta
Premisa: Alguien las mató
Conclusión: Hay un asesinato

Es decir, encuentran los cadáveres y **suponen que alguien las mató,** por lo tanto: hay un asesinato; y si hay un asesinato, hay un asesino. Vamos a ver cómo empiezan a construir las hipótesis que les ayuden a encontrar al criminal.

Construcción de hipótesis

Una vez descubierto el crimen, el forense levantó un informe donde se relata la manera cómo fue encontrado cada cadáver, la hora de su muerte, así como también se encontraron los diarios de los huéspedes asesinados. Esta información sobre el informe forense y diarios personales es deducida del diálogo de los detectives. La autora no se detiene a describir esos detalles**.**

Los investigadores comienzan por preguntarse si hay alguna pista dada por el informe del forense. Dice Maine que no hay indicios y pasa a describir cómo muere cada uno de los diez visitantes a la isla. Sin embargo, esta descripción arroja una información muy valiosa que los lleva a ambos a formular una hipótesis preliminar.

Los huéspedes pasan varios días en la isla desierta; alguien los proveía de alimentación y alguien los transportó; ese alguien sabía que estaban allí: suposición: ese alguien (Isaac Morris) puede ser el asesino. Esta hipótesis preliminar, como tal, está elaborada a título de ensayo, y, conjuntamente, con los hechos observados ¿construirá? la solución.

Los hechos, como la muerte de Morris, hacen que los investigadores descarten la suposición de la autoría de éste; sin embargo, indudablemente es un cómplice. De todas maneras, la complicidad de Morris no *parece* conducirlos a nada importante. Al descartarlo como posible autor de los crímenes y también como figura de indicios relevantes, dirigen su mirada hacia las gentes del pueblo cercano; sin embargo, hay una explicación para que no maliciaran nada y eso los "libra" de sospechas a su vez; fueron prevenidos sobre la presencia de extraños; se les dijo que esas personas vivían una especie de prueba de resistencia para vivir en un lugar deshabitado.

Eso "explicaba" el porqué no se sintieron intrigados por la llegada de los excepcionales huéspedes.

Ahora bien, alguien descubrió los cadáveres. ¿Cómo los descubrió? ¿Quién lo hizo? ¿Por qué fue a la isla?

> - Fred Narracot – continuó Maine -, el que llevó a los invitados a la isla, me hizo una observación muy significativa. Se extrañó de la clase de invitados de Mr. Owen. "Nada que ver con la clase de amigos del joven Robson". Creo que fue por verlos tan tranquilos y normales por lo que, a pesar de las órdenes de Morris, se fue a la isla en cuanto oyó hablar de las señales de SOS.
> - ¿Cuándo llegaron Narracot y los otros hombres en su socorro?
> - Las señales fueron captadas el día once por la mañana por un grupo de *boy scouts*. Ese día fue imposible llegar a la isla por el mal estado del mar. Sólo se pudo abordar en la tarde del día doce. Todos afirman que nadie pudo salir de la isla antes de su llegada. Había una marejada tremenda después de la tormenta.
> (A. Christie, 2004, 176. Cursivas en el texto).

El inspector investigó a cada invitado y halló que cada uno tenía una deuda pendiente con la justicia ordinaria. De una u otra manera eran culpables de un delito y no habían sido juzgados. Sir Thomas Legge se impacienta después de oír la narración sobre quién era quién y exclama:

> - Toda esta historia es absurda, increíble…Es inadmisible que diez personas sean asesinadas en una roca en medio del mar…y que ignoremos quién ha cometido el crimen, por qué y cómo – gritó
> Maine carraspeó y dijo:
> - Permítame contradecirle, señor, sobre lo del móvil. Sabemos **por qué** más o menos. Algún loco con una idea particular de la justicia. Salió a la búsqueda de personas a las que la justicia ordinaria no pudo castigar. Escogió a diez. Que fuesen inocentes o culpables poco importa.
> (A. Christie, 2004, 173. Cursivas en el texto).

"La mente activa ve problemas donde el tonto sólo ve objetos familiares" (I. Copi, 1976, 495). El inspector capta unos indicios que le pasan inadvertidos al subjefe: la deuda de cada huésped con la justicia. Así elabora su hipótesis "novedosa" para explicar el suceso:

el asesino se toma la justicia por su propia mano. Queda un problema por resolver: ¿Cómo sale de la isla?

> - Nuestro maniático reunió diez personas…digamos condenadas a muerte. Fueron ejecutadas por un tal U.N. Owen, quien cumplió su tarea y se evaporó como el humo.
> - Eso sería un truco de magia prodigioso, Maine – observó Legge -. Pero seguramente tiene otra explicación.
> - Usted sabe, señor, que si ese hombre se encontraba en la isla, no habría podido abandonarla y, siguiendo las notas escritas por los interesados, este Mr. Owen no puso jamás los pies en la Isla del Negro. Por tanto, sólo quedaría una solución viable: **¡Que Owen fuese uno de los diez!** (A. Christie, 2004, 176. Negritas nuestras).

El inspector aventura una hipótesis que, al contrastarla con otros datos sacados de los diarios personales de algunos de los huéspedes, lo llevan erróneamente a descartarla para elaborar otra: **alguien más estaba en la isla y ése fue el asesino**. Esta conclusión los encierra en un círculo que les impide establecer el **quién**.

Resumamos brevemente qué averiguan:

Hay un asesinato en "serie", es decir, el **qué**.

Hay un móvil, "hacer justicia" que responderían a la pregunta ¿**por qué**?

Logran averiguar los medios mediante los cuáles se cometió cada acto criminal. Éstos serán el ¿**cómo**?

A su vez, estos medios esbozan, de una u otra forma, el momento, el contexto, es decir, el ¿**cuándo**?

Sin embargo, no logran determinar el ¿**quién**?

Habla el asesino

En el "Documento anexo" a la novela, el asesino comienza por describir sus propias características, es decir, analiza su personalidad para mostrar cómo siente placer al ver morir o dar él mismo la muerte a un ser vivo; pero a su vez, con un profundo sentido de la justicia. Detalles muy importantes, porque esos rasgos que podrían aparecer como insignificantes o secundarios, unidos a otras circunstancias, señalan claramente al asesino.

Describe a cada víctima (culpable de algún hecho delictivo y no juzgado) e idea juntarlos en una isla deshabitada y tomar la justicia por su propia mano. Así queda claro el **porqué** y el **dónde**. Para definir el **cómo**, se remite a una vieja canción infantil llamada "Ten Little Niggers" y, por ello, escoge a diez víctimas; la décima es el "cómplice" que busca para llevar a los invitados y dotarlos de víveres. No descuida pormenores, y si no escogía a Morris como la décima víctima, le fallaría la canción. Cada verso indica quién muere y cómo. Planifica prolijamente su "muerte" para alejar las sospechas, y, al final, se suicida en la isla, de tal manera que a la policía le es imposible identificar el **quién**.

Razonamiento del detective

Es significativo señalar que previo a todas las hipótesis, se ha presumido que el asesino era alguien que NO estaba en la isla.

Todos los huéspedes pasan días en la isla desierta
Alguien los llevó y los proveyó de víveres
Ese alguien es Isaac Morris
Puede ser el caso de que "Morris sea el asesino"

Esta hipótesis es rechazada, porque Morris también muere *asesinado*, pero no en la isla. Es importante enfatizar el asesinato del sospechoso, pues éste, a su vez, estaba en deuda con la justicia: había traficado con drogas, y resultaba culpable de la muerte de una joven, quien se había suicidado presumiblemente por el abuso de las drogas. Este último detalle, pasa desapercibido, o no es analizado en profundidad por el detective. Dirige su mirada a la gente del pueblo:

Todos los huéspedes mueren asesinados
La gente del pueblo sabía que ellos estaban allí sin sospechar nada sobre su presencia
Puede ser el caso de que "alguien del pueblo sea el asesino".

Hipótesis descartada. Un dato adicional les hace ver que no hay *vínculos* que los incrimine: Fueron advertidos sobre una supuesta *prueba de resistencia* en la isla. Sin embargo, *alguien encontró los cadáveres*. Aparece otra persona que transportó huéspedes y fue uno de los que llegaron a la isla advertidos por unos boys scouts. Este dato

adicional arroja una pista: *nadie pudo salir de la isla ni llegar a ésta, debido a la fuerte marejada que azotó la zona.*

Agreguemos que al estudiar el perfil de cada huésped salta el móvil: todos tenían deuda pendiente con la justicia y en consecuencia: *el asesino se toma la justicia por su propia mano*. El detective, con los datos e hipótesis determina que: Por tanto, sólo quedaría una solución viable: *¡Que Owen fuese uno de los diez!*

A pesar de dar en el blanco, descarta esa solución, porque el orden en el que se suceden las muertes, según las notas y diarios de los huéspedes, sólo pueden señalar como asesino a Armstrong o a Vera Claythorne, y las evidencias de sus muertes no dejan duda alguna: fueron asesinados; luego,…alguien más debía estar en la isla. Sigue la incógnita: ¿**Quién fue**?

Esquema resumido de los investigadores

Problema que se debe resolver: Hay un crimen en serie (diez huéspedes en una isla desierta)

Hipótesis preliminar: *El asesino no es uno de los diez huéspedes*

Todos los huéspedes pasan días en la isla desierta

Isaac Morris los llevó y los proveyó de víveres

Puede ser el caso que *Morris sea el asesino* (Segunda hipótesis, totalmente coherente con la primera)

Morris es asesinado

Morris no es el asesino

Se presume que *el asesino esté en el pueblo* (Tercera hipótesis, coherente con la primera)

Nadie del pueblo va a la isla ni sospecha nada

Alguien descubre los cadáveres

Se presume que *ese alguien haya llegado antes de las muertes en serie y sea el asesino* (Cuarta hipótesis, coherente con la primera)

Hay una fuerte marejada que impide al acceso a la isla

Nadie llega a la isla antes de las muertes

Queda una incógnita en este momento sin resolver por el detective

Todos los huéspedes están en deuda con la justicia
Las muertes las idea alguien con un loco sentido de la justicia
Se presume que *el asesino se toma la justicia por su propia mano* (Quinta hipótesis, coherente con la primera)

Todos los huéspedes fueron asesinados
El asesino no puede abandonar la isla
Se presume que *el asesino era uno de los huéspedes de la isla (Solución del problema)*

Contrastación de la solución

A pesar de dar en el blanco, descarta esa solución, porque el orden en el que se suceden las muertes, según las notas y diarios de los huéspedes, sólo pueden señalar como asesino a Armstrong o a Vera Claythorne, y las evidencias de sus muertes no dejan duda alguna que fueron asesinados; luego,…alguien más debía estar en la isla. Sigue la incógnita: **¿Quién fue?**

Por qué el detective no da con la solución

Hay unos datos que no podían dejarse de lado, so pena de errar la solución, como efectivamente pasó.

La letra de la canción: En ella se da el orden estricto en el que suceden las muertes.

La muerte de Morris, aunque no sucedió en la isla, estaba vinculada con Owen. También tenía deudas con la justicia. Es la muerte número 11.

El informe del forense: detalla la hora de cada muerte.

Si a cada "descubrimiento" del detective, éste hubiera sumado alguno de estos detalles, no hubiera descartado la hipótesis correcta.

Veamos: La muerte del último, según el documento enviado por el asesino, es el propio asesino. Con el informe del forense se podía determinar que no fue Vera Claythorne la última en ser ajusticiada; fue el juez Margrave, y éste aparece, según la letra de la canción, en el quinto lugar. Luego, ¡hay una contradicción evidentísima!: según una

interpretación es el quinto; según el documento es el último. Con este dato hubiese bastado para corroborar la hipótesis: ¡Que Owen fuese uno de los diez!

Si se hubiese agregado el dato de la muerte de Morris, no desvinculada de Owen, la cuenta daba once muertes y no diez, lo que añadido a los datos les hubiera conducido a lo explicado en líneas precedentes.

A la luz de la abducción

Sin entrar en largas explicaciones sobre la abducción en este escenario, recordemos que la validez de un determinado argumento dependerá, para Peirce, esencialmente de dos principios: la certeza y la fecundidad para ofrecer mayores evidencias; reservando para la abducción la capacidad para "descubrir"; mientras la abducción es parte del desarrollo de ese descubrimiento, la inducción los "prueba" (cf. CP 8.384, 8.387s, 1913).

> La diferencia específica entre la inducción y la abducción es aquí que la abducción forma parte del proceso de descubrimiento, mientras que la inducción forma parte del proceso de probar los descubrimientos. Mediante la inducción un general dado será sólo confirmado o falsado por experimentos futuros.
> (M. Hoffmann, http://www.unav.es/gep/AN/Hoffmann.html)

La aplicación del "razonar" abductivo no puede partir sino de la distinción hecha por el propio Peirce y tan atinadamente comentada por M. Beuchot en **Analogía Filosófica**:

> Peirce distingue el razonar *hacia* una hipótesis del razonar *desde* una hipótesis. Justamente la *abducción* es el razonamiento *hacia* la hipótesis, esto es, desde los hechos hacia la hipótesis que les señala su causa o los explica (…) En la abducción pasamos de la observación de ciertos casos a la suposición de un principio general que dé cuenta de ellos. Se

> pasa del efecto a la causa, esto es, se explica. De hecho, es el único razonamiento que introduce nuevas ideas[3].

Para Peirce la abducción no es un procedimiento caprichoso de preferencia por una hipótesis, sino que es fundamental tener presente lo indispensable que será delimitar los contextos donde se elaborarán las hipótesis. Es decir, que "una inferencia abductiva es 'lógica' , si y sólo si el conjunto elegido de posibles hipótesis está determinado por un cierto conjunto de contextos que son asumidos como relevantes en una situación histórica determinada" (M. Hoffmann, http://www.unav.es/gep/AN/Hoffmann.html), y esos contextos, en el caso de la novela analizada, están determinados por las condiciones de aislamiento de los huéspedes (imposibilidad de salir de la isla por las condiciones atmosféricas) y la cercanía del pueblo, por tan sólo señalar algunos de ellos. Tampoco olvidemos que la hipótesis abductiva es aquélla que nos ofrece la mejor de las soluciones posibles.

En el caso que nos ocupa de la novela policial de Agatha Christie, extraordinario ejemplo práctico del "razonar abductivo", podemos explicar los argumentos de la siguiente manera:

Los hechos de la soledad de los huéspedes y su clara e inobjetable "necesidad" de alimentarse hacen ver a los investigadores que alguien debió proveerlos de víveres; es así que este proveedor era el único que entró en contacto con ellos, de tal manera que "ese alguien fue el asesino"; como se sabía quién era ese alguien, se construyó una primera hipótesis (abductiva): Se presume que *Morris sea el asesino* a partir de*: Hecho: Isaac Morris los llevó y los proveyó de víveres.* Repitamos que la hipótesis abductiva es aquélla que ofrece la "mejor" solución.

Pero, nuevos hechos observados descartan esta hipótesis: *Morris es asesinado*, luego, no es el asesino. *Hay gente en el pueblo que curiosamente parece no sorprenderse por la presencia de huéspedes*

[3] Mauricio Beuchot: "Abducción y analogía" en *C. S. Peirce y la Abduccción* en **Analogía Filosófica XII/1**, (1998), 1-187. Cursiva mías. En este número monográfico, hay una excelente selección de artículos sobre Peirce presentados en **VI Congreso Internacional de la International Association for Semiotic Studies**, celebrado en Guadalajara, México, 1997, que me han servido de telón de fondo para la explicación del argumento central de la novela analizada.

en la isla. Eso los hace sospechosos, hechos que conducen hacia otra hipótesis (abductiva) *el asesino está en el pueblo.*

Nuevos hechos, la descartan; y surgen hechos, como que *alguien, obviamente, descubrió los cadáveres*, que apuntan a otra hipótesis (abductiva): puede ser que *ese alguien haya llegado antes de las muertes en serie y sea el asesino.*

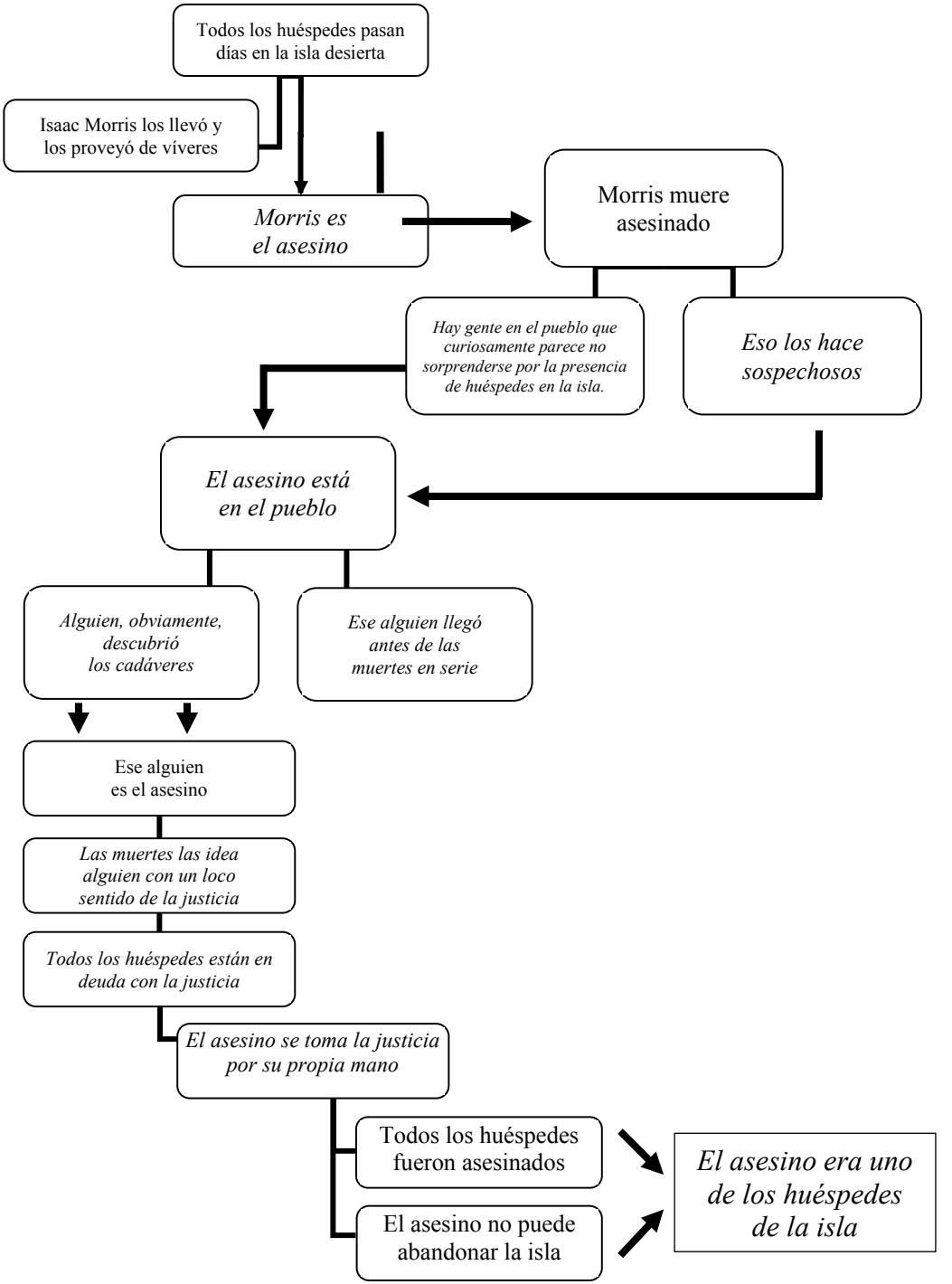
Hechos como que *Todos los huéspedes están en deuda con la justicia* conducen a suponer que *las muertes las idea alguien con un loco sentido de la justicia* y ésta a su vez a *se presume que* (abductiva) *el asesino se toma la justicia por su propia mano*.

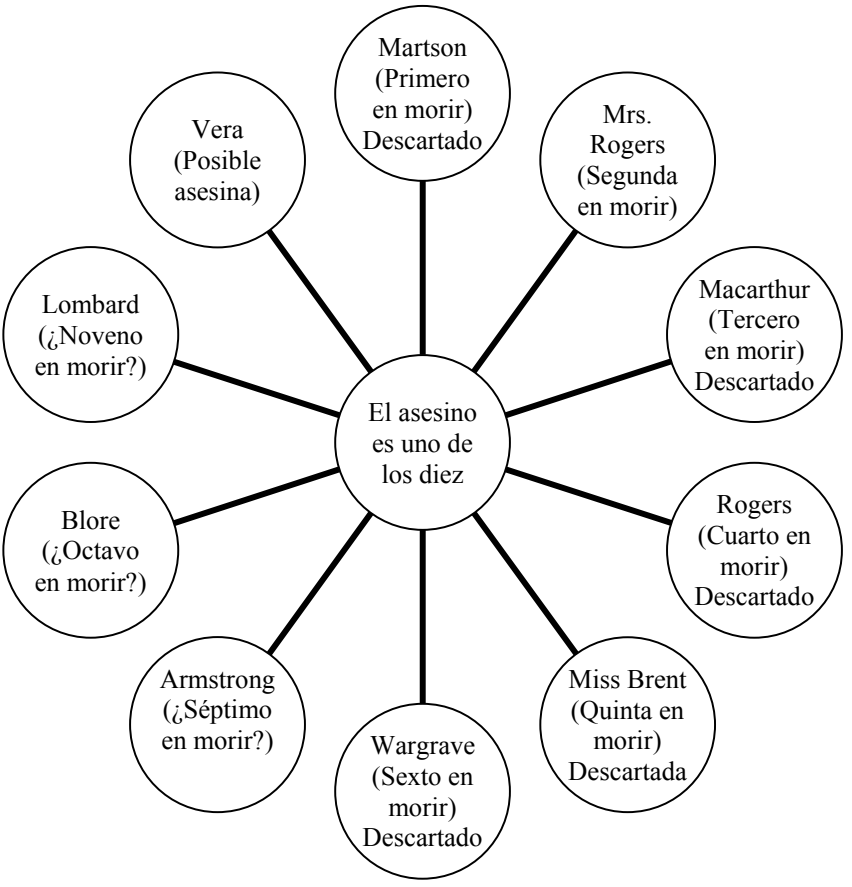
Cada hecho e hipótesis anteriores, resumidos en que *todos los huéspedes fueron asesinados* y que *el asesino no puede abandonar la isla* conducen a *se presume que el asesino era uno de los huéspedes de la isla (Solución del problema. Abductiva).*

¿Quién es?

El asesino es uno de los diez. Hipótesis: *Es el último en morir*. Hechos prueban que no fue la última persona, pues se ha supuesto que la última fue Vera, por un dato extraído de la observación de una silla, y de nuevo vuelven a la primera hipótesis *el asesino no era uno de los diez (abductiva):*

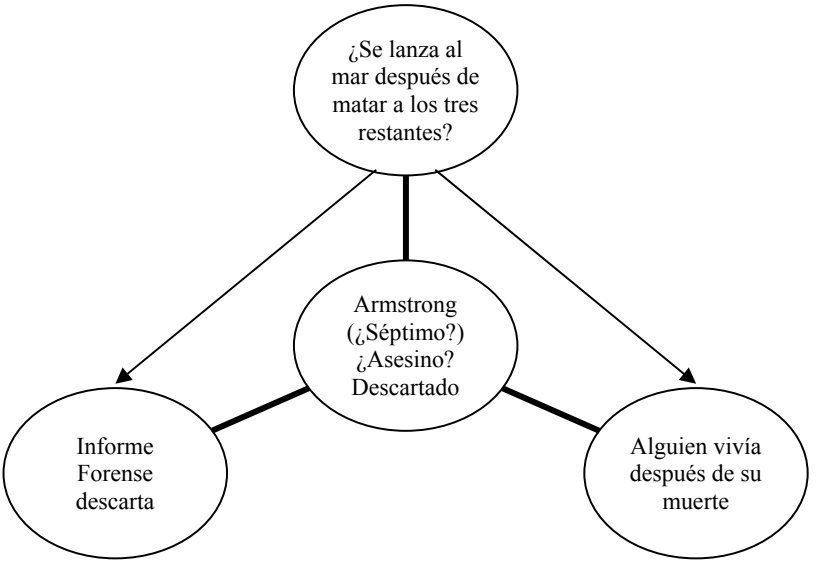
> - Adivino lo que quiere decir, señor. Que fue Vera quien mató a Lombard, se llevó el revólver a la casa, tiró sobre Blore el bloque de mármol y después se colgó.
> - Esta suposición sería admisible hasta cierto punto En su cuarto, sobre una silla, se encuentran las mismas huellas de algas que en sus zapatos, lo que prueba que se subió sobre la silla, pasó la cuerda alrededor de su cuello y tiró la silla de un puntapié.
> Pero fíjese, señor. *La silla no estaba caída en el suelo, sino como las demás, junto a la pared*. Luego, fue puesta en su sitio después de la muerte de Vera Claythorne *por alguien*.
> (Cursiva en el texto)

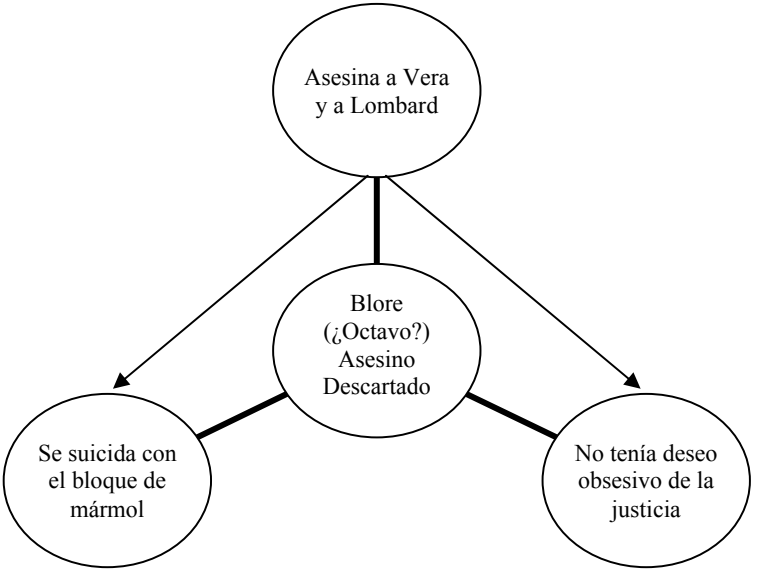

Todos los huéspedes pasan días en la isla desierta
Isaac Morris los llevó y los proveyó de víveres
Morris es el asesino
Morris muere asesinado
Hay gente en el pueblo que curiosamente parece no sorprenderse por la presencia de huéspedes en la isla.
Eso los hace sospechosos
El asesino está en el pueblo
Alguien, obviamente, descubrió los cadáveres
Ese alguien llegó antes de las muertes en serie
Ese alguien es el asesino
Las muertes las idea alguien con un loco sentido de la justicia
Todos los huéspedes están en deuda con la justicia
El asesino se toma la justicia por su propia mano
Todos los huéspedes fueron asesinados
El asesino no puede abandonar la isla
El asesino era uno de los huéspedes de la isla

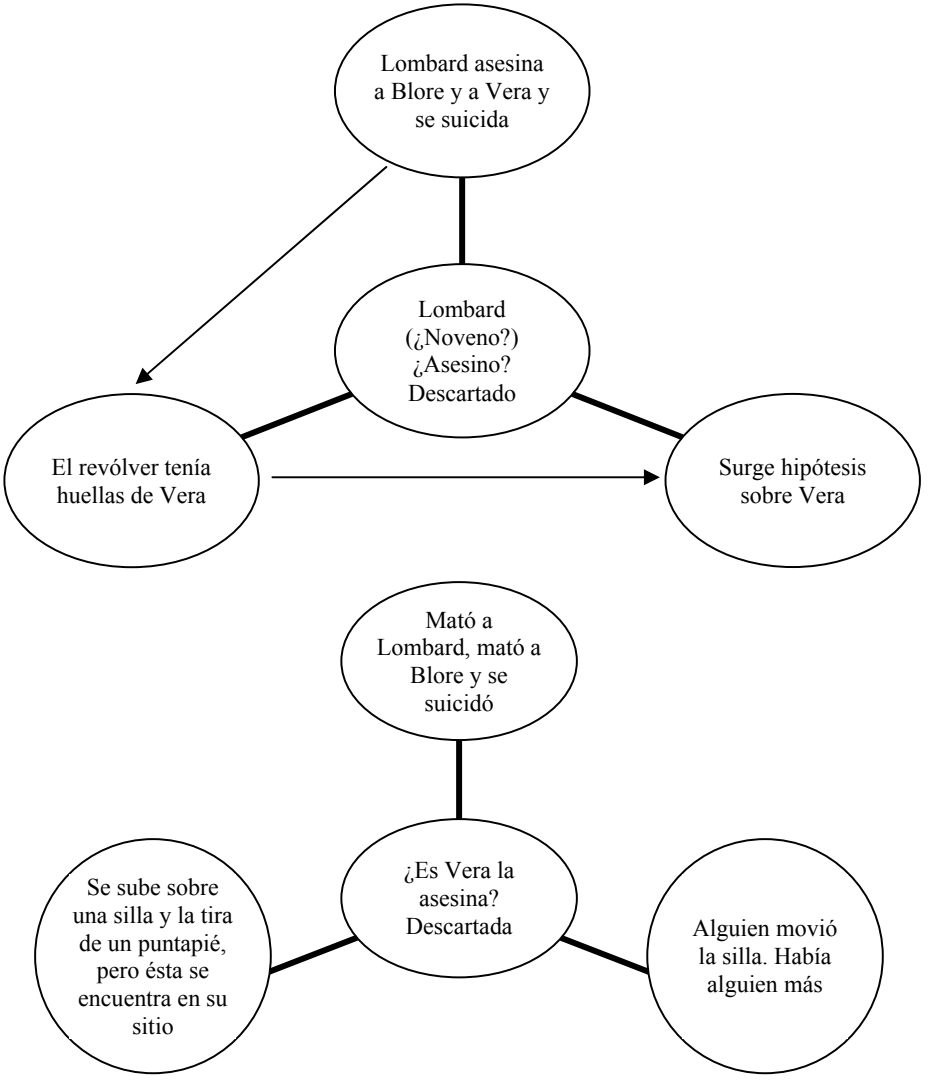


Al quedar cuatro huéspedes vivos, cada uno se vuelve altamente sospechoso. Hay que suponer a cada uno como el asesino y ver qué pasa con esa hipótesis.

Sin embargo, si el análisis efectuado a los últimos cuatro huéspedes se le hubiese realizado a quien era lógico suponer con un sentido obsesivo de la justicia, la hora de la muerte no hubiera coincidido según el informe del forense y se hubiese podido establecer quién era el asesino.


¿Se lanza al mar después de matar a los tres restantes?
Armstrong (¿Séptimo?) ¿Asesino? Descartado
Informe Forense descarta
Alguien vivía después de su muerte


Asesina a Vera y a Lombard
Blore (¿Octavo?) Asesino Descartado
Se suicida con el bloque de mármol
No tenía deseo obsesivo de la justicia



¿Logró usted descifrar el enigma?
¿Quién fue el asesino?

Si el análisis efectuado a los últimos cuatro huéspedes se le hubiese realizado a quien era lógico suponer con un sentido obsesivo de la justicia, la hora de la muerte no hubiera coincidido según el informe del forense y se hubiese podido establecer quién era el asesino: el Juez Margrave.



Centro de Filosofia das Ciências
da Universidade
de Lisboa

Breve Histórico sobre o Centro de Filosofia das Ciências da Universidade de Lisboa

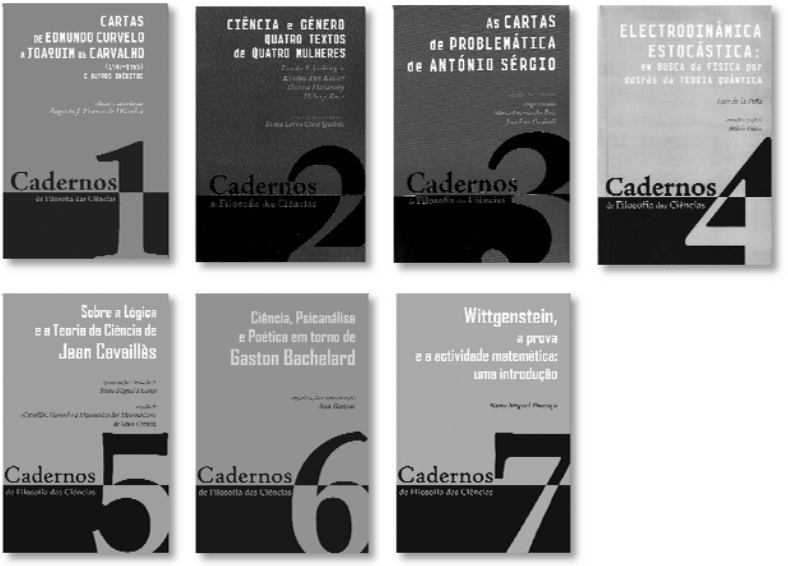
O Centro de Filosofia das Ciências da Universidade de Lisboa (CFCUL) iniciou oficialmente as suas actividades em Julho de 2003 e conta com uma equipa interdisciplinar que inclui membros e colaboradores, nacionais e internacionais, de filiações institucionais distintas. O CFCUL conta ainda com o apoio de prestigiados consultores: Andrade e Silva (Faculdade de Ciências da Universidade de Lisboa), Franco Selleri (Universidade de Bari, Itália), Daniel Andler (Sorbonne, Paris IV, França) e Maria José Frapolli (Universidade de Granada, Espanha).

O CFCUL, que desenvolve as suas actividades no interior de sete linhas de investigação, pretende promover o desenvolvimento da Filosofia das Ciências em Portugal através do estímulo e apoio à investigação, da participação em projectos de investigação nacionais e internacionais, da colaboração com grupos de trabalho igualmente nacionais e internacionais, do apoio a estudos pós-graduados e da ligação da investigação à formação que se desenvolve em torno do "Mestrado em História e Filosofia das Ciências da FCUL".

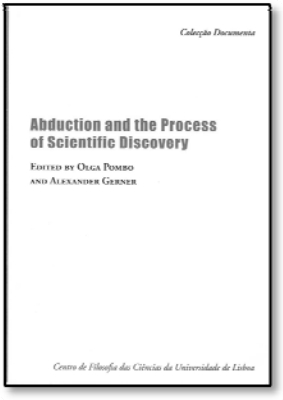
O CFCUL tem ainda como objectivo a difusão de conhecimentos na área da Filosofia das Ciências promovendo colóquios nacionais e internacionais, conferências, workshops e outros encontros, levando a cabo um "Seminário Permanente de Filosofia das Ciências", publicando trabalhos decorrentes da investigação realizada, nomeadamente, livros, artigos, traduções e editando a revista "Conceito" e quatro colecções: "Cadernos de Filosofia das Ciências", "Thesis", "Documenta" e "Fundamentos e Desafios do Evolucionismo".

Outras Publicações:

- *Cadernos de Filosofia das Ciências:*



- *Colecção Documenta:*



- *Colecção Thesis:*



Este Livro foi impresso em Lisboa,
em Janeiro de 2009, pela Guide Artes Gráficas Lda.
Miolo: 304 páginas / papel IOR 80g / 16x23cm.
Capa: plastificação brilho / cartolina chromo 240g.
Acabamento: cadernos fresados
e colados a quente, com vinco à francesa.
Principais Letras utilizadas:
Times New Roman / Palatino Linotype / Garamond