Boletín de la Sociedad de Lógica , Metodología y Filosofía de la Ciencia

ISSN: 1577-2292
SLMFCE

Especial Octubre de 2016

Contenido:

Editan:

Inmaculada Perdomo, Antonio Blanco y Concha Martínez

III Conferencia de Graduados de la Sociedad de Lógica, Metodología y Filosofía de la Ciencia en España

Valencia, 1 y 2 de Junio de 2016


III CONGRESO DE GRADUADOS DE LA SOCIEDAD DE LÓGICA, METODOLOGÍA Y FILOSOFÍA DE LA CIENCIA EN ESPAÑA
1 y 2 de junio de 2016
Facultad de Filosofía y CC. de la Educación
Avda. Blasco Ibáñez 30, Valencia
Ponente invitado:
SAMIR OKASHA
(University of Bristol, Reino Unido)
www.solofici.org
Sociedad de Lógica, Metodología y Filosofía de la Ciencia
VNIVERSITAT DE VALÈNCIA
Facultat de Filosofia i Ciències de l'Educació

III Congreso de Graduados de la SLMFCE

III Congreso Graduados área de Lógica y Filosofía de la Ciencia (Valencia, 1-2 junio 2016) Salón de Grados, Facultad de Filosofía y CC. Educación, Univ. de València

Programa

Miércoles 1 junio

12:00 – 12:15 Presentación del Congreso

12:15 – 13:15 "Cognición Ad Hoc y Estructura Conceptual: Dos Nociones de Concepto en el Marco de la Teoría de Prototipos" J. Vicente Hernández Conde (Univ. País Vasco). Comentarista: Josep Corbí (Univ. València)

13:30 – 14:30 "Supervaluations and Horwich's Fixed-Point Theory of Truth" Sergi Oms (Univ. Barcelona). Comentaristas: Concha Martínez y Matteo Plebani (Univ. Santiago de Compostela)

----- Comida ------

16:00 – 17:00 Conferencia Samir Okasha (Bristol University, Reino Unido): "Darwin's Argument Reconsidered"

17:15 – 18:15 "Symbiosis research or why present models of natural selection fail" Javier Suárez (Exeter Univ., Reino Unido) – Comentarista: Cristian Saborido (UNED, Madrid)

18:30 – 19:30 "Selección natural y complejidad" Giorgio Airoldi (UNED, Madrid) – Comentarista: Diego Rasskin-Gutman (Instituto Cavanilles de Diversidad y Biología Evolutiva, Univ. València)

19:45 – 20:45 Asamblea SLMFCE

Jueves 2 de junio

9:00 – 10:00 "Entrenching the epistemological side of computer simulations: explanation and unification" Juan M. Durán (Universität Stuttgart). Comentarista: María Caamaño (Univ. Valladolid)

10:15 – 11:15 "Epistemological disjunctivism as a solution for underdetermination-based skepticism" Eduardo Martínez Zoroa (Univ. Barcelona). Comentarista: Tobies Grimaltos (Univ. València)

11:30 – 12:30 "Relevance Theory, grammar and processing effort: how grammar diversity affects the explicitness of utterances" Joan Gimeno (Univ. València). Comentarista: Antonio Blanco (Univ. Complutense Madrid)

12:45 – 13:45 Conferencia Samir Okasha (Bristol University, Reino Unido): "Rationality from a Biological Perspective"

--------- Comida ---------

15:30 – 16:30 "The causal structure of Evolutionary Theory: the scope and limits of the force interpretation" Víctor Luque (Univ. València). Comentarista: José A. Diez (Univ. Barcelona)

16:45 – 17:45 "Polisemia e infraespecificación semántica." Marina Ortega (Univ. País Vasco). Comentarista: Cristina Corredor (Univ. Valladolid)

18:00 – 19:00 "How to fix what is said" Claudia Picazo (Univ. Barcelona). Comentarista: Jordi Valor (Univ. València)

Clausura del Congreso

El congreso está financiado por la Sociedad de Lógica, Filosofía y Metodología de la Ciencia en España y por la Facultad de Filosofía y CC. Educación (Univ. de València).

Crónica de la III Conferencia de Graduados de la SLMFCE

La III Conferencia de Graduados de la Sociedad de Lógica, Metodología y Filosofía de la Ciencia de España (SLMFCE) tuvo lugar en Valencia los días 1 y 2 de junio de 2016. Se desarrolló en el Salón de Grados de la Facultad de Filosofía y Ciencias de la Educación de la Universitat de València y fue financiada por la SLMFCE y la Facultad de Filosofía y Ciencias de la Educación de la Universitat de València.

La tercera edición de la Conferencia de Graduados de la SLMFCE fue la edición con más participantes hasta la fecha. Fueron recibidas un total de 27 propuestas, de las cuales se aceptaron 10 (el 37%). Hubo pluralidad temática, viéndose presentes en la conferencia las áreas (señalo entre corchetes el número total de comunicaciones por área) de Filosofía de la Biología [3], Filosofía del Lenguaje [3], Epistemología y Filosofía de la Mente [2], Filosofía General de la Ciencia [1] y Filosofía de la Lógica
[1].

La Conferencia contó con la presencia del ponente invitado Samir Okasha (Bristol University, Reino Unido), cuyo trabajo se ha centrado en la Filosofía de la Biología, la Epistemología y la Teoría de la Decisión. Entre sus méritos destaca la obtención en 2009 del prestigioso premio Imre Lakatos por su monografía Evolution and Levels of Selection (2006). Samir Okasha ofreció dos conferencias, tituladas "Darwin's Argument Reconsidered" y "Rationality from a Biological Perspective".

En la apertura de la III Conferencia de Graduados, a cargo de la presidenta de la sociedad, Concha Martínez, y del representante del comité organizador local, Valeriano Iranzo, se remarcó el objetivo fundamental de este evento: favorecer que los jóvenes investigadores y recientes doctorados del área den a conocer su trabajo. La estructura de cada sesión fue la acostumbrada en ediciones anteriores: cada comunicación de un joven investigador va seguida de un comentario a la misma por parte de un experto sénior en el área, con lo que se pretende orientar y enriquecer la discusión posterior. Por otro lado, la conferencia es una ocasión ideal para que los investigadores de distintas universidades, tanto españolas como extranjeras, establezcan vínculos.

La primera de las comunicaciones fue llevada a cabo por José Vicente Hernández- Conde (UPV/EHU), y versó sobre cuestiones del área de Epistemología y Filosofía de la Mente. La comunicación, titulada "Cognición Ad Hoc y estructura conceptual: dos nociones de concepto en el marco de la teoría de prototipos", empezó con la exposición de las principales concepciones de concepto: la invariantista (los conceptos se identifican con cuerpos de conocimiento estable entre individuos y tiempos) y la contextualista (muchos conceptos dependen del contexto). A continuación, el autor presentó la nueva propuesta del marco de la cognición ad hoc, desarrollada por Daniel Casasanto y Gary Lupyan, según la cual no hay conceptos independientes del contexto y todos los conceptos son ad hoc. Desde esta perspectiva, cada instanciación de un concepto sería específicamente producida en cada momento, a partir de la información contextual disponible para el sujeto en ese momento. Con respecto a la relación entre esta propuesta y la estructura de los conceptos, el autor defendió la tesis de que una teoría de prototipos articulada mediante un modelo dimensional basado en espacios de similitud conceptual es la mejor manera de dar cuenta de la estructura y la instanciación de los conceptos en el marco de la cognición ad hoc. Este trabajo fue comentado por Josep E. Corbí (Univ. València), quien planteó las diferentes dificultades de mantener una posición contextualista respecto de muchos de nuestros conceptos.

A continuación, Sergi Oms (UB) presentó su comunicación "Supervaluations and Horwich's Fixed-Point Theory of Truth", la cual se inscribió en el área de la Filosofía de la Lógica. En su presentación expuso la teoría minimalista de la verdad de Paul Horwich, la cual consiste en todas las instanciaciones del esquema T "<p> es verdadera sii p" aplicado a proposiciones. El problema de esta propuesta es que la presencia en la teoría minimalista de la verdad de proposiciones que afirman su propia falsedad la convertiría en inconsistente en el marco de la lógica clásica. Para evitar esta inconsistencia y no tener que renunciar a la lógica clásica, Horwich restringió las instanciaciones del esquema T que componen su teoría de la verdad. Para ello, apeló a una construcción similar a la propuesta por Saul Kripke en "Outline of a Theory of Truth" (1975) basada en el concepto de «fundamentación» [groundedness].

A lo largo de su exposición Sergi Oms estableció y evaluó la estructura formal de esta cons-

Crónica de la III Conferencia de Graduados de la SLMFCE

trucción planteada por Horwich. Como comentarista, Concha Martínez (Univ. Santiago de Compostela) señaló que podría argumentarse que ciertas restricciones planteadas en la propuesta de Horwich hacen que sea problemático entenderla como una propuesta minimalista.

La sesión vespertina de la primera jornada se inició con la primera de las ponencias ofrecidas por Samir Okasha. En esta conferencia, titulada "Darwin's Argument Reconsidered", Okasha analizó desde una perspectiva moderna el estatus de uno de los argumentos ofrecidos por Charles Darwin en El origen de las especies. Se planteó la cuestión de si, tal y como consideró Darwin, la teoría de la selección natural puede explicar por qué encontramos organismos tan bien adaptados. La selección natural no hace altamente esperable la adaptación, ya que en general no le concede una alta probabilidad; y por lo tanto no satisface uno de los requisitos que Carl G. Hempel consideraba necesarios para dar cuenta de un explanandum. Pero, como han señalado diferentes autores por medio de contraejemplos, hacer esperable el explanandum no es realmente una condición necesaria para la explicación. La argumentación desarrollada en la conferencia culminó en la conclusión de que, a pesar de no hacerla esperable, la selección natural es la mejor explicación que tenemos de la adaptación.

A continuación, se presentaron dos comunicaciones que trataban cuestiones relativas a la Filosofía de la Biología. En "Symbiosis research or why present models of natural selection fail", Javier Suárez (University of Exeter, Reino Unido) presentó el problema filosófico en torno a la noción de simbiosis. Por lo general, los biólogos entienden por simbiosis cualquier tipo de interacción biológica entre organismos de diferentes especies. Javier Suárez ofreció una caracterización restringida de la misma, entendiendo por simbiosis aquellas en las que hay adquisición de un organismo por parte de otro, y como, consecuencia de la interacción durante un largo periodo, surgen nuevas estructuras y rutas reproductivas/metabólicas (las cuales no hubieran surgido de otro modo), haciendo que la relación sea necesaria para al menos uno de los organismos. Respecto de este tipo de relaciones de simbiosis, centrales para la biología evolutiva, el autor planteó que los dos principales modelos actuales de Selección Natural (la "concepción heredada" y el modelo basado en interactores y replicadores) no pueden dar cuenta ni de su importancia, ni de su rol en la evolución. Cristian Saborido (UNED, Madrid) llevó a cabo el comentario de esta comunicación y formuló diversas cuestiones acerca de las relaciones de simbiogénesis y la necesidad de formular un modelo de Selección Natural que evidencie su relevancia.

Giorgio Airoldi (UNED, Madrid) presentó su trabajo "Selección natural y complejidad". Aunque hay consenso respecto a que el mecanismo de la Selección Natural puede dar cuenta de la variedad de organismos, no lo hay respecto a si puede explicar la complejidad de los organismos. Para abordar esta segunda cuestión, Giorgio Airoldi parte de la clasificación de hechos evolutivos propuesta por Massimo Pigliucci, quien distingue entre Heritability (variación de frecuencias genéticas), Evolvability (hechos evolutivos que dependen de la arquitectura genética y de constricciones del desarrollo, y afectan a la adaptación a largo plazo y a la exploración del espacio fenotípico más allá de la mera recombinación de rasgos actuales) e Innovation (hechos evolutivos que conllevan la superación de constricciones genéticas y de desarrollo, y producen novedades fenotípicas relevantes). Las concepciones unidimensionales de la complejidad, aquellas que solo atienden a la optimización de la eficacia, solo pueden dar cuenta del incremento de la complejidad en los hechos evolutivos del nivel Heritability. En su intervención, el autor expuso la tesis de que para medir adecuadamente la complejidad en todos los niveles de hechos evolutivos, no basta con fijarse en la optimización de la eficacia (no es suficiente con apelar a la Selección Natural), sino que también es necesario prestar atención a la robustez (capacidad de un sistema para mantener sus funciones frente a perturbaciones internas y externas). El comentarista, Diego Rasskin-Gutman (Instituto Cavanilles de Diversidad y Biología Evolutiva, Univ. València), señaló la relevancia de tener en cuenta la capacidad evolutiva a la hora de medir la complejidad y la importancia de clarificar la relación entre este rasgo y la eficacia y la robustez..

La primera jornada de la III Conferencia finalizó con la celebración de la asamblea de la Sociedad de Lógica, Metodología y Filosofía de la Ciencia de España, a la cual asistieron tanto miembros de la junta directiva como socios de la SLMFCE.

La segunda jornada se inició con la presentación de la comunicación "Entrenching the epistemological side of computer simulatios: explanation and unification",

Crónica de la III Conferencia de Graduados de la SLMFCE

centrada en temas propios de la Filosofía General de la Ciencia, a cargo de Juan M. Durán (High- Performance Computing Center Stuttgart, Universität Stuttgart). El autor defendió la posibilidad de entender las simulaciones computacionales como explicaciones. Para ello, apeló a la concepción unificacionista de la explicación desarrollada por Philip Kitcher y planteó que las simulaciones pueden entenderse como explicaciones de tipo unificacionista en las que el explanans está constituido por el modelo de simulación mismo y el explanandum por el resultado de la simulación computacional. Aunque señaló que para poder sostener adecuadamente esta interpretación es necesario adaptar a las simulaciones la noción de comprensión como unificación de una pluralidad de fenómenos. El autor indicó que una limitación de su propuesta es que, aunque se ajusta a las simulaciones computacionales basadas en ecuaciones, no puede dar cuenta de otros tipos de simulaciones computacionales (como las basadas en agentes) con metodologías distintas. Este trabajo fue comentado por María Caamaño (Univ.Valladolid), quien planteó la importancia de establecer una comparativa entre esta propuesta y la interpretación de las simulaciones computacionales como experimentos.

Seguidamente tuvo lugar la charla de Eduardo Martínez Zoroa (UB), titulada "Epistemological disjunctivism as a solution for underdetermination-based skepticism". En esta comunicación, Eduardo Martínez abordó las paradojas y los argumentos escépticos basados en el principio de subdeterminación. En ellos se apela a escenarios en los cuales el agente, a pesar de ser masivamente engañado, tiene experiencias subjetivamente indistinguibles de aquellas que tiene ordinariamente, para señalar la falta de justificación de muchas de nuestras creencias. Centrando el análisis en el escenario clásico del cerebro en la cubeta, el autor planteó que el disyuntivismo epistemológico puede ser utilizado para resolver las paradojas escépticas basadas en el principio de subdeterminación y refutar los argumentos escépticos basados en este principio. La idea básica del disyuntivismo epistemológico es que en un caso óptimo de percepción el agente tiene un respaldo epistémico fáctico (tener respaldo epistémico perceptivo de p implica la verdad de p) y reflexivamente accesible (uno puede saber que obtiene este respaldo epistémico por la mera reflexión). Esta consideración permite distinguir los casos ordinarios de percepción (los cuales son óptimos) de los casos de mala percepción o no percepción; ya que en ellos, a pesar de que son subjetivamente indistinguibles de la percepción ordinaria, no se da este tipo de respaldo y no son óptimos. Tobies Grimaltos (Univ. València), quien llevó a cabo el comentario de esta comunicación, conectó la propuesta de Eduardo Martínez con otras paradojas escépticas similares a la del cerebro en la cubeta y señaló dificultades adicionales para las posturas escépticas radicales.

La tercera intervención de la jornada corrió a cargo de Joan Gimeno Simó (Univ. València), con la comunicación "Relevance Theory, grammar and processing effort: how grammar diversity affects the explicitness of utterances". La teoría de la relevancia se basa en el Principio de Relevancia, según el cual todo acto ostensivo de comunicación comunica la presunción de su propia relevancia óptima. Un estímulo es óptimamente relevante si conduce al oyente a inferir el número máximo de asunciones relevantes con el menor esfuerzo de procesamiento. Los teóricos de la relevancia sostienen que todo hablante que quiera comunicar un conjunto de asunciones por medio de una declaración codificará solo la información mínima requerida para que el oyente lo entienda. Según esta teoría, la relevancia y el esfuerzo de procesamiento son los dos principales elementos de los que depende la explicitación de un enunciado. En su comunicación, Joan Gimeno cuestionó esta consideración y planteó que en muchas ocasiones la gramática de la lengua empleada juega un papel más importante a la hora de determinar el grado de explicitación adecuado. Para respaldar su tesis apeló a ejemplos de diferentes lenguas (español, inglés, chino, italiano,…), evidenciando así que no se trata de un fenómeno aislado de una gramática en concreto. El comentario corrió a cargo de Antonio Blanco (Univ. Complutense de Madrid), quien destacó la originalidad de la propuesta y planteó la conveniencia de conectarla con otras investigaciones acerca de la diversidad y la relatividad lingüísticas.

La sesión matutina de la segunda jornada finalizó con la segunda de las conferencias ofrecidas por Samir Okasha, titulada "Rationality from a Biological Perspective". Versó sobre la relación entre el concepto de adaptación y el de racionalidad. Al respecto, Okasha defendió que entender la adaptación como protorracionalidad es, a pesar de sus limitaciones, una idea viable. Apeló a la clasificación de Alex Kacelnik, quién distingue tres tipos de ra-

Crónica de la III Conferencia de Graduados de la SLMFCE

cionalidad: filosófica y psicológica (PP-racionalidad), biológica (B-racionalidad) y económica (E-racionalidad). La ponencia concluyó con la idea de que la interpretación de la adaptación como protorracionalidad, vinculada a la noción de B-racionalidad, permite dar cuenta del uso de terminología intencional en la conducta ecológica y de por qué los modelos de elección racional tienen aplicaciones biológicas.

La tarde del segundo día comenzó con la comunicación "The causal structure of Evolutionary Theory: the scope and limits of the force interpretation" de Víctor J. Luque (Univ. València). En esta comunicación, centrada en el campo de la Filosofía de la Biología, el autor abordó la estructura causal de la teoría de la evolución. Expuso la analogía de fuerzas planteada respecto de la teoría de la evolución y el debate en torno a cómo encajar la deriva en ella. A este respecto, señaló las dificultades de entender la deriva como una condición de fondo y la conveniencia de entenderla como un factor evolutivo. Finalmente, defendió que para establecer adecuadamente la analogía de fuerzas respecto de la teoría de la evolución es necesario entender las fuerzas evolutivas como difference-makers e introducir entre las condiciones de fondo lo que él denominó el "Principio de Estasis". Éste principio fue definido de la siguiente manera: un sistema evolutivo en el que no hay ningún difference-maker (selección, deriva, migración, mutación,…) permanecerá en estasis (no experimentará ningún cambio). En su comentario José A. Díez (UB) señaló la conveniencia de desligar la definición del Principio de Estasis del concepto de factor evolutivo entendido como difference-maker, para evitar problemas de analiticidad.

Las dos últimas comunicaciones se ocuparon de la Filosofía del Lenguaje. En la primera de ellas, "Polisemia e infraespecificación semántica", Marina Ortega Andrés (UPV/EHU) presentó la diferencia entre polisemia (los significados de las diferentes palabras están relacionados) y homonimia (los significados de las diferentes palabras no están relacionados), y expuso las diferentes explicaciones de la polisemia defendidas en la actualidad: la tesis de la enumeración de sentidos, la teoría del léxico generativo y la teoría de la relevancia. A continuación, tomando como punto de partida la teoría del léxico generativo (la polisemia se forma a través de mecanismos generativos que ocurren dentro del léxico a partir del significado más común del término a otros sentidos que se superponen) desarrollada por James Pustejovsky, defendió la consideración de los conceptos como complejos (que contienen los distintos sentidos de la palabra polisémica), generativos e influenciados por el contexto y la información extralingüística del mundo. Cristina Corredor (Univ. Valladolid), en su detallado comentario, planteó cuestiones relativas a los ejemplos utilizados por Marina Ortega (principalmente el verbo inglés bake) y señaló la importancia actual de la teoría del léxico generativo.

En último lugar, Claudia Picazo Jaque (UB) presentó una comunicación titulada "How to fix what is said". La autora expuso los diferentes tipos de propuestas que buscan determinar qué es lo que se ha dicho: las centradas en el hablante (el contenido de una declaración depende de lo que el hablante intenta comunicar), las independientes de los interlocutores (el contenido de una declaración depende del significado lingüístico junto con algunos parámetros objetivos del contexto de uso) y las centradas en el intérprete (el contenido de una declaración depende de cómo es plausible o razonable interpretarla). Posteriormente señaló las dificultades que encuentran estas perspectivas en su aplicación, y planteó la conveniencia de un modelo mixto basado en la idea de que las condiciones de satisfacción de un predicado están determinadas por la actividad en juego. Algunas de las ventajas de este modelo son que la distinción entre lo que el hablante dice y lo que quiere decir no es borrosa, y que lo que es dicho coincide con una interpretación razonable, la cual hace responsables a los hablantes y justifica ciertos cursos de acción. El comentario de esta comunicación corrió a cargo de Jordi Valor (Univ. València), quien planteó la posibilidad de responder a las dificultades de las actuales propuestas sobre cómo determinar lo que se ha dicho, por medio de la construcción de un modelo pluralista que las articule y aplique en cada caso una de ellas.

En la clausura de la Conferencia, la presidenta de la sociedad y el representante del comité organizador local agradecieron el trabajo de los diferentes participantes en la conferencia y subrayaron la relevancia de eventos como este, dirigidos a jóvenes investigadores.

Saúl Pérez-González

Universitat de València

III Conferencia de Graduados de la SLMFCE

Cognición *Ad Hoc* y Estructura Conceptual: Dos Nociones de Concepto en el Marco de la Teoría de Prototipos

José V. Hernández-Conde
Universidad del País Vasco
jhercon@gmail.com

RESUMEN: Recientemente Casasanto y Lupyan (2015) han sostenido que no hay conceptos independientes del contexto: todo concepto sería construido *ad hoc* en el momento de su instanciación. En este artículo muestro que el marco de la cognición *ad hoc* puede caracterizarse mediante una teoría de espacios de similaridad conceptual, y distingo dos nociones de concepto asociadas a diferentes etapas de su ciclo de vida (almacenamiento e instanciación). Este modelo reúne virtudes de enfoques antagónicos: (a) invariantista: la estabilidad del con-cepto almacenado permitiría registrar nueva información; y (b) contextualista: la dependencia contextual del concepto instanciado explicaría nuestra capacidad de adaptación.

PALABRAS CLAVE: Conceptos, Cognición *ad hoc*, Contexto, Teoría de prototipos

1. Introducción

Los conceptos juegan un papel fundamental en procesos cognitivos tales como categorización, inferencia, aprendizaje, memoria, etc. Por lo general se los identifica con cuerpos de conocimiento sobre los miembros de una cierta categoría, razón por la cual la noción de *concepto* es clave en las teorías del conocimiento y el comportamiento, para explicar cómo los sujetos clasifican distintos objetos, y realizan generalizaciones. En este trabajo mi punto de partida será que los conceptos son herramientas cognitivas empleadas por nuestras mentes en categorizaciones, pues cualquier otra tarea cognitiva (inferencias, resolución de problemas, toma de decisiones, etc.) está sustentada por unas categorizaciones previas.

Mi presente discusión se centra en el grado de dependencia contextual que puede atribuirse a nuestros conceptos. Con tal propósito presentaré primero las posturas contextualista e invariantista, como principales propuestas en cuanto a ese grado de dependencia contextual. Más tarde introduciré el marco de la cognición *ad hoc* propuesto por Casasanto y Lupyan (C&L), en torno a la tesis de que todos los conceptos dependen del contexto, así como cuál es su principal limitación, a saber, su no-articulación en torno a una teoría concreta sobre la estructura de los conceptos. Tras ello, argumentaré a favor de la posibilidad de articular el marco de la cognición *ad hoc* mediante una teoría de prototipos caracterizada en términos de espacios de similaridad conceptual, e identificaré cuatro posibles fuentes de dependencia contextual. Sobre la base de esa propuesta distinguiré dos nociones de concepto, que identificaré con etapas diferentes de su ciclo de vida: (1) *conceptos almacenados*, o información almacenada por nuestro sistema cognitivo de modo persistente; (2) *conceptos instanciados*, resultantes de procesos cognitivos tales como categorización, inferencia, etc. Esta propuesta reúne muchas de las virtudes de los enfoques invariantista y contextualista, dejando de lado la cuestión de cómo es posible la mutua comprensión de los mensajes intercambiados por los integrantes de una conversación.

2. Conceptos y dependencia contextual: *invariantismo vs. contextualismo*

La visión tradicional identifica los conceptos con cuerpos de conocimiento estables entre individuos y tiempos. Esta concepción *invariantista* (Machery 2009), permite explicar tanto la acumulación de conocimiento por parte los individuos, como su capacidad para comunicarse con otros sujetos:

> -*Si los conceptos no fuesen estables para un mismo sujeto S*, entonces no habría nada que proporcionase la continuidad que necesitaría un concepto *C* para almacenar nueva información sobre él (pues no habría modo de reconocer nuevas instancias de *C*).
>
> -*Si los conceptos no fuesen estables y compartidos entre los interlocutores de una conversación*, entonces la mutua comprensión de los mensajes intercambiados no sería posible (pues el oyente podría interpretar un término *T* de modo distinto a lo significado por el hablante).

Otros autores sostienen que muchos conceptos dependen del contexto, en el sentido de que son constructos creados al vuelo de modo específico en cada ocasión (Barsalou 1993; Sperber y Wilson 1995; Carston 2002; Prinz 2002; Malt 2010). La con-cepción *contextualista* permite explicar la adaptación del comportamiento ante entornos cambiantes.

3. Marco de la cognición *ad hoc*

Recientemente, C&L (2015) han propuesto una atractiva tesis: *no hay conceptos independientes del contexto*, esto es, todos los conceptos son conceptos *ad hoc*. De ser así, cada instanciación de un concepto sería generada al vuelo, de manera específica para cada ocasión en función de

III Conferencia de Graduados de la SLMFCE

la información contextual disponible en ese momento.

Estos autores argumentan que la aparente estabilidad de los conceptos es sólo debida a los rasgos compartidos por sus diferentes instanciaciones, pero que en realidad no hay nada invariante a todos ellos, en línea con la conclusión de Wittgenstein (1953: §66-100) en su discusión de los parecidos de familia para el término "juego" (contraria a la existencia de un conjunto de propiedades comunes a las entidades referidas con ese término).

C&L sostienen que, siendo el estado cognitivo del sujeto parte de su contexto, y estando el cerebro en continuo cambio, eso implica que los conceptos de los sujetos son intrínsecamente variables. Por ello afirman que los conceptos sólo existen en el preciso instante en que son instanciados, es decir, cuando son empleados por un sujeto para categorizar, comunicarse, realizar inferencias, etc.: "Concepts are not something we *have* in the mind, but something we *do with* the mind." (C&L 2015: 546)

No obstante, C&L centran su trabajo en la cuestión de la instanciación de conceptos, dejando de lado el problema de qué estructuras cognitivas pueden sostener esas instanciaciones. Sin embargo, para poder aceptar el marco de la cognición *ad hoc* ambas cuestiones (instanciación conceptual y estructura conceptual) requieren la misma atención, por lo que una detallada explicación de esa estructura conceptual resulta necesaria.

En la última parte de este trabajo muestro que los procesos cognitivos en los que reconocemos conceptos (*instanciación conceptual*) y la información almacenada que dichos procesos instancian (*almacenamiento conceptual*) pueden verse como dos etapas distintas del ciclo de vida de un concepto. Sin embargo, antes de abordar esa cuestión veremos cómo el marco de la cognición *ad hoc* puede articularse mediante la teoría de prototipos.

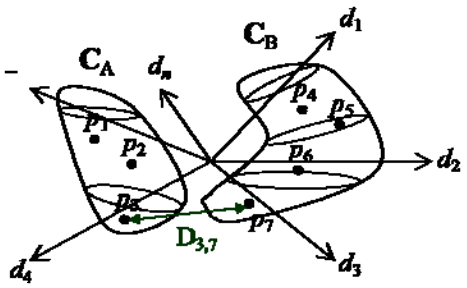
4. Teoría de prototipos (y *espacios de similaridad conceptual*)

La teoría de prototipos nace con el propósito de explicar los fenómenos de tipicalidad identificados en muchos conceptos (Rosch y Mervis 1975; Rosch 1978), algo que no era posible con la teoría clásica (en la cual los conceptos se definían en términos de condiciones necesarias y suficientes). Para la teoría de prototipos los conceptos son prototipos, esto es, representaciones cuya estructura codifica información estadística sobre las propiedades que suelen tener los miembros de su categoría. En su versión más general (*modelos dimensionales*) algo será clasificado bajo un concepto si posee en cierto grado un número suficiente de las propiedades asociadas a ese concepto. Por tanto, la pertenencia de un objeto a un cierto concepto se determina en función de la similaridad existente entre ese objeto y el prototipo asociado al concepto en cuestión (la cual se determina en virtud de las propiedades compartidas por ambos).

Si los objetos y los prototipos de los conceptos se localizan en un espacio geométrico cuyas dimensiones fuesen las propiedades constitutivas de los conceptos relevantes (en el contexto considerado), estaríamos ante una *teoría de espacios de similaridad conceptual* (Churchland 1989; Gärdenfors 2000).

Teorías de espacios de similaridad conceptual

Las teorías de espacios de similaridad conceptual conciben a la mente como un hiperespacio representacional donde las *dimensiones* representan cómo los objetos pueden diferir, los *puntos* representan objetos, las *regiones* representan conceptos, y las *distancias* son inversamente proporcionales a la *similaridad* entre objetos y conceptos (Gauker 2007).



Por tanto, un objeto será categorizado bajo un concepto si y sólo si sus valores en cada dimensión producen una *n*-tupla que cae dentro de la región asociada a ese concepto.

Los prototipos asociados a los conceptos resultan de generalizar las propiedades de los objetos escogidos como miembros tentativos de cada categoría, mediante un proceso de maximización de semejanzas (o minimización de distancias), entre los objetos evaluados y los prototipos provisionales. Las fronteras de las regiones conceptuales podrían resultar de un teselado de Voronoi del hiperespacio conceptual que tomase como entrada los prototipos de los conceptos relevantes.

Distinción entre prototipos y regiones conceptuales

Ahora bien, aunque prototipos y regiones conceptuales (cuando estas regiones conceptuales proceden de un teselado de Voronoi) son nociones interdefinibles (pues dado cualquiera de ellos puede determinarse unívocamente el otro), mi tesis es que la *información almacenada* por nuestro sistema cognitivo de los conceptos son sus prototipos, y no sus regiones y/o fronteras asociadas, puesto que:

-Lo que resulta de generalizar ejemplares de un concepto es un prototipo, no una región.

-Para categorizar sólo se necesitan son los prototipos de los

Cognición Ad Hoc y Estructura Conceptual

conceptos relevantes.

-Requiere menos memoria almacenar un prototipo que una región conceptual.

COGNICIÓN *AD HOC* EN LA TEORÍA DE PROTOTIPOS

Veamos ahora cómo tendría lugar la instanciación de un concepto en una teoría de espacios de similaridad conceptual. En estas teorías la similaridad es inversamente proporcional a la distancia entre los objetos (y/o prototipos), la cual podría estar dada por la distancia Minkowski entre dos objetos (y/o conceptos) *A* y *B*, en donde $x_i^{[Y]}$ representase el valor de la i-ésima dimensión asociada al objeto/concepto *Y*:

$$d(A,B) = \left(\sum_{i=1}^{n} w_i \left| X_i^{[A]} - X_i^{[B]} \right|^p \right)^{1/p}$$

El parámetro *p* determina el tipo de métrica (si p=1 la métrica es *city-block*; si p=2 la métrica es *euclidiana*). Además, puesto que cada dimensión podrá contribuir de modo distinto a la similaridad, los pesos w_i darán cuenta de esa posible distinta contribución.

La anterior expresión se corresponde con las distancias Minkowski ordinarias. Sin embargo, esas distancias podrán ser distintamente ponderadas conforme a diferentes criterios. Por ejemplo, los pesos podrían ser función del número de ejemplares en base a los que se determinó el prototipo asociado a cada concepto. En tal caso, la distancia-de-comparación entre un objeto *O* y un concepto C_k (con prototipo p_{ck}), referida como $d_{ck}(O, P_{Ck})$, podría expresarse bajo un esquema de ponderación multiplicativa (Okabe *et al.* 1992: 119-134):

$$d_{C_k}(O, P_{C_k}) = u_k \, d(O, P_{C_k})$$

Por tanto, la categorización de un objeto *O* bajo uno u otro concepto tiene lugar mediante un proceso cognitivo que evalúa las distancias de *O* con respecto a los prototipos de todos los conceptos relevantes (en un contexto dado), tras lo cual *O* se clasifica bajo el concepto más próximo (es decir, más semejante a *O*). Es en procesos cognitivos como el anterior en donde tiene lugar la *instanciación* de los conceptos, una instanciación que suele consistir en la evaluación de la similaridad de un objeto frente a un grupo de conceptos relevantes.

En un modelo como éste existirían al menos cuatro factores que pueden hacer que la instanciación de los conceptos dependa del contexto, a saber: (i) conceptos relevantes, (ii) tipo de métrica *p*, (iii) importancia de las dimensiones w_i, y (iv) ponderación de los conceptos u_k. La razón es que cualquier variación en alguno de estos factores contextuales producirá un teselado distinto del hiperespacio conceptual y, consecuentemente, instanciaciones distintas de los conceptos considerados. De este modo, la teoría de prototipos (concebida como un espacio de similaridad conceptual) puede dar perfecta cuenta de la tesis principal defendida por C&L, a saber, que todos los conceptos son *ad hoc* (es decir, que la instanciación de todo concepto depende del contexto en el que esa instanciación ocurre).

5. Ciclo de vida de un concepto: *almacenamiento vs. instanciación*

Mi tesis final es que deben diferenciarse dos nociones de concepto: (a) *concepto almacenado* o información registrada por nuestro sistema cognitivo con respecto a una cierta categoría; y (b) *concepto instanciado*, asociado con la instanciación de parte de la información almacenada para el caso de un contexto concreto.

CONCEPTOS ALMACENADOS

En una teoría de prototipos articulada en torno a espacios de similaridad conceptual la única información que precisa ser almacenada por nuestro sistema cognitivo es la localización de los prototipos asociados a cada concepto. Esa información es todo lo que se necesita para que un concepto pueda instanciarse en un contexto particular (esto es, para determinar las distancias y similaridades entre ese concepto y cualquier objeto), y su registro proporciona la continuidad necesaria para acumular nueva información sobre ese concepto con el paso del tiempo.

CONCEPTOS INSTANCIADOS

Sin embargo, la mera información almacenada sobre un concepto (esto es, la localización de su prototipo) no basta para explicar cómo dicho concepto se aplica en tareas tales como categorización, inferenciación, etc. La razón de ello es que en dichas tareas cognitivas lo que interviene no es el concepto *como almacenamiento*, sino la instanciación de esa información almacenada (la cual dependerá del contexto y, por consiguiente, también de la información almacenada sobre otros conceptos). Esto es lo que se ha llamado hasta ahora instanciación de un concepto. Los *conceptos instanciados* pueden identificarse con los conceptos *ad hoc* postulados por C&L.

En línea con lo sostenido por C&L, los conceptos instanciados se forman al vuelo cada vez que un objeto precisa ser categorizado, en función de cuál sea el contexto. Por ello se puede decir que un concepto instanciado no es una entidad psicológica, sino el mero resultado de un proceso, que sólo existiría en el preciso momento en que el proceso de instanciación concluye (esto es, cuando termina de evaluarse la similaridad con respecto a los conceptos relevantes). Por consiguiente, cabe decir que los conceptos instanciados no son algo que exista sino algo que sucede al final de cualquiera de los procesos cognitivos mencionados.

En resumen, los *conceptos instanciados* son algo que ocurre,

Cognición Ad Hoc y Estructura Conceptual

esto es, el resultado de un proceso cognitivo en donde parte de la información almacenada sobre esa categoría se emplea (conjuntamente con información almacenada sobre otras categorías, así como con otro tipo de información dependiente del contexto) para decidir la categorización (o no) de un objeto concreto bajo un cierto concepto.

DOS FASES EN EL CICLO DE VIDA DE UN CONCEPTO

En consecuencia, *almacenamiento* e *instanciación* no se corresponderían con nociones asociadas a teorías alternativas de lo que un concepto es, sino que explicarían dos fases distintas en el ciclo de vida de un concepto:

-Primeramente, cuando el concepto se adquiere nuestro sistema cognitivo almacena cierta información sobre él, lo que dentro de una teoría de espacios de similaridad conceptual sería la localización de su prototipo asociado. Esa información es el *concepto almacenado* que, estando registrada bajo una misma entrada mental, es lo que dota de continuidad al concepto en cuestión, y lo que explica nuestra capacidad para acumular nueva información sobre un mismo concepto.

-No obstante, el mero concepto almacenado no determina la categorización o no de un objeto bajo un cierto concepto, pues existen otros factores de dependencia contextual: los conceptos relevantes, el tipo de métrica, la importancia de las dimensiones y el peso dado a los conceptos. Es por ello que decimos que la noción que interviene en los procesos de categorización, inferencia y demás no es la de concepto almacenado, sino la de *concepto instanciado*.

Este ciclo de vida de un concepto no sería lineal sino circular, y en él la información almacenada permanecería estable bajo la forma de un concepto almacenado hasta la siguiente ocasión en que fuera necesario su empleo en una tarea de categorización, momento en el cual el concepto sería nuevamente instanciado. Luego, como resultado de la categorización de un nuevo objeto bajo ese concepto, el prototipo asociado a este último podría ser revisado la siguiente vez que los procesos de ajuste conceptual fuesen ejecutados, lo cual podría dar lugar a una actualización de la información asociada al concepto almacenado. En definitiva, *almacenamiento* e *instanciación* no serían más que las dos caras de una misma moneda que, en último término, es el concepto.

6. Conclusiones

En estas páginas he mostrado cómo el marco de la cognición *ad hoc* de C&L puede caracterizarse mediante una teoría de prototipos articulada mediante espacios de similaridad conceptual. Esa caracterización sería compatible con la tesis de que no hay conceptos independientes del contexto (esto es, con que todos los conceptos son conceptos *ad hoc*), en la medida en que hemos identificado cuatro posibles fuentes de dependencia contextual: conceptos relevantes, tipo de métrica, importancia de las dimensiones y pesos de los conceptos.

Sobre la base de esta propuesta, se han distinguido dos nociones de concepto (almacenado e instanciado), como dos etapas diferenciadas de su ciclo de vida. La primera, concepto *como almacenamiento*, se identificó con la información que el sistema cognitivo registra sobre los conceptos, cuya propiedad principal es su persistencia en el tiempo. La segunda, o concepto *como instanciación*, sólo ocurriría en el momento final de sus procesos cognitivos asociados, a pesar de lo cual es la responsable de la manifestación externa de dichos conceptos.

La ventaja de esta aproximación es que reúne virtudes de los enfoques contextualista e invariantista. En cuanto al contextualismo, articula de manera satisfactoria un marco, el de la cognición *ad hoc* de C&L, compatible con las evidencias existentes en contra de definiciones y/o núcleos conceptuales, y a favor de que muchos conceptos son constructos creados al vuelo para cada ocasión, lo que explicaría nuestra capacidad de adaptarnos ante entornos cambiantes. Con respecto al invariantismo, mi propuesta explica cómo, a pesar de la absoluta dependencia contextual de los conceptos (instanciados), los conceptos (almacenados) gozan de la estabilidad necesaria para acumular nueva información sobre ellos.

Agradecimientos: La investigación que condujo al desarrollo del presente trabajo fue financiada por una beca predoctoral FPI de la Universidad del País Vasco, y llevada a cabo dentro del proyecto de investigación Lenguaje y Pensamiento: Significado Lingüístico y Conceptos (FFI2014-52196-P), subvencionado por el Ministerio de Economía y Competitividad.

Referencias:

BARSALOU, L.W. (1993): "Flexibility, structure, and linguistic vagary in concepts: Manifestations of a compositional system of perceptual symbols", en A.F. Collins, S.E. Gathercole, M.A. Conway & P.E. Morris (eds.), *Theories of Memory*, Lawrence Erlbaum Associates, Hillsdale, pp. 29-101.

CARSTON, R. (2002): *Thoughts and Utterances*, Blackwell, London.

CASASANTO, D.; LUPYAN, G. (2015): "All concepts are ad-hoc concepts", en E. Margolis & S. Laurence (eds.), *The Conceptual Mind: New Directions in the Study of Concepts*, MIT Press, Cambridge MA, pp. 543-566.

III Conferencia de Graduados de la SLMFCE

CHURCHLAND, P.M. (1989): "On the nature of theories: A neurocomputational perspective", en *Minnesota Studies in the Philosophy of Science* 14, pp. 59-101.

GÄRDENFORS, P. (2000). *Conceptual Spaces: The Geometry of Thought*, MIT Press, Cambridge MA.

GAUKER, C. (2007): "A critique of the similarity space theory of concepts", *Mind & Language* 22(4), pp. 317-45.

MACHERY, E. (2009): *Doing Without Concepts*, Oxford University Press, Oxford.

MALT, B.C. (2010): "Why should we do without concepts", *Mind & Language* 25(5), pp. 622-633.

OKABE, A.; BOOTS, B.; SUGIHARA, K.; CHIU, S.N. (1992). *Spatial Tessellations: Concepts and Applications of Voronoi Diagrams*, John Wiley & Sons, New York.

PRINZ, J. (2002): *Furnishing the Mind*, MIT Press, Cambridge MA.

ROSCH, E. (1978): "Principles of categorization", en E. Rosch & B. Lloyd (eds.), *Cognition and Categorization*, Erlbaum Associates, Hillsdale, pp. 27-48.

ROSCH, E.; MERVIS, C.B. (1975): "Family resemblances: Studies in the internal structure of cate-gories", *Cognitive Psychology* 7(4), pp. 573-605.

SPERBER, D.; WILSON, D. (1995): *Relevance: Communication and Cognition* (2nd ed.), Blackwell, Oxford.

WITTGENSTEIN, L. (1953): *Philosophical Investigations*, G.E.M. Anscombe & R. Rhees (eds.), G.E.M. Anscombe (trans.), Blackwell, Oxford.

III CONFERENCIA DE GRADUADOS DE LA SLMFCE

Supervaluations and Horwich's Fixed-Point Theory of Truth

Sergi Oms (UB, Logos)

Abstract. Horwich's theory of truth, Minimalism, is inconsistent in classical logic due to the Liar paradox. Horwich has tried to overcome this difficulty by restricting the instances of the T-schema that constitute the minimalist theory of truth so that no paradox can be formulated. In this paper I make precise Horwich's attempt to give a fixed point construction to specify which instances of the T-schema are to be included in the theory and I will show that the fixed point exists and it is consistent. As a mater of fact, it is identical to Kripke's fixed point using the Supervaluational scheme. Finally, some unsatisfactory properties are noted and a temptative solution is suggested.

I

Horwich's theory of truth, called 'Minimalism', consists of all the instances of the T-schema applied to propositions:

(T-schema) < p > is true iff p

Horwich (1998, 2001, 2010b) has presented and defended Minimalism. Now, as it is well known, the proposition that asserts its own non-truth (let's call it 'the Liar') makes the theory consisting of just all instances of the T-schema inconsistent with classical logic. Until recently, Horwich's response to this problem had been very succinct. In his (1998) he claims that the lesson the Liar tells us is that not all the instances of the T-schema are to be included as axioms in the theory (Horwich (1998, p. 42)). Thus, the minimalist theory of truth consists of a restricted collection of instances of the T-schema; only those that do not engender Liar-like paradoxes.

Thus, Horwich's strategy in front of the Liar consists of restricting the instances of the T-schema that constitute the minimalist theory of truth so that no paradox can be formulated; what I called before 'the paradoxical instances of the T-schema' must be ruled out of the truth theory. Then, though, a natural question arises: which instances of the T-schema are to count as paradoxical? Horwich (1998, p. 42) proposes two conditions that this restriction should meet:

Maximality Instances of the T-schema cannot be excluded unnecessarily; the minimal theory of truth should be, if possible, a maximal consistent collection of instances of the T-schema.

Specification There must be a constructive specification of the instances of the T-schema excluded from the minimal theory of truth. Such specification should be as simple as possible.

Horwich has offered, in his (2010a, p. 90), a construction which, although not being maximal, would follow a constructive specification of which instances of the T-schema are in the theory.

Horwich wants to use a construction similar to the one proposed in Kripke (1975) and uses the notion of groundedness so that the grounded sentences are the ones whose instances of the T-schema constitute the min- imalist theory of truth. The construction, though, is presented in a very loose way and it si needed of clarification. My intention in this paper is to make it precise and have a deeper look at the results.

Horwich wants to use a construction similar to the one proposed in Kripke (1975) and takes the grounded sentences to be the ones whose instances of the T-schema constitute the minimalist theory of truth. This already raises some doubts about whether a deflationist can use the notion of groundedness in order to specify its theory of truth. Let us think of this construction, hence, as a mere technicality.

For perspicuity, let's suppose we have a classical first-order language L and an expanded language L+ = L ∪ {Tr} with a truth predicate Tr and suppose, furthermore, that for every formula $\varphi \in$ L+ we can express its canonical name $\langle\varphi\rangle$ in L via some codification. I will suppose that L is strong enough to prove the Diagonal Lemma.

Given a model for the base language, N with domain D, I will use $\langle$N , $A\rangle$ to refer to the model of the expanded language L+ whose interpretation of Tr is A, which will be a set of codes of formulas of L+. I will use $|\alpha|M = 1$ to mean that the formula α has semantic value 1 in the model M (and the same for having semantic value 0). Given a set of formulas Γ, I will use $|\Gamma|M = 1$ to mean that, for every $\gamma\in\Gamma$, $|\gamma|M = 1$.

SENT will be the set of (codes of) sentences of L+; as I said I am supposing that, via some suitable codification, *SENT* $\subseteq D$.
Let's begin with the construction. It will consist of a series $H\sigma$ of sets of sentences of L+ defined for every ordinal σ and relative to a model N for the base language. We need, first, the following definitions.

Definition Let's define the following.
For any set A of formulas of L+, $A^- = \{\varphi \in L+ : \neg\varphi \in A\}$.

Supervaluations and Horwich's Fixed-Point Theory of Truth

For any $\varphi \in$ L+, $T\varphi$ is the φ-instance of the T-schema, i.e. $Tr\langle\varphi\rangle \leftrightarrow \varphi$.
For any set A of sentences of L+, $TA = \{T\varphi : \varphi \in A$ or $\varphi \in A^-\}$.

Now, Horwich presents a construction involving a single truth predicate and a series of sets of sentences of L+ (which he calls 'languages') which we could try to characterize in the following way, given a model N for the base language and for any ordinal σ,

$H0 = \{\varphi \in L: |\varphi|N = 1\}$
$H\sigma+1 = \{\varphi \in L+: H\sigma \cup TH\sigma \models \varphi\}$
$H\lambda = \cup_{i<\lambda} Hi$

where λ is a limit ordinal.

Horwich claims that 'our language L is the limit of the expanding sublanguages'(Horwich (2010a, p. 90)); he means with that that the formulas in the alleged limit of the sequence are the formulas whose instances of the T-schema constitute the minimalist theory of truth. It is a good guess to suppose that what Horwich has in mind is something similar to what Kripke (1975) presents in his construction; that is, a fixed point of the construction. Hence, we are looking for an ordinal τ such that $H\tau = H\tau+1$.
First, if we want to show the existence of a fixed point, we must prove that the series is monotone (in this long abstract I skip all the proofs).

Lemma 1.1 (Monotonicity) *If $\tau \leq \rho$, then $H\tau \subseteq H\rho$.*

Theorem 1.2 (Fixed point) *There is an ordinal τ such that $H\tau = H\tau+1$.*

I will call the fixed point of the construction ***H***. Thus, Horwich's theory of truth, the Minimalist theory of truth, is ***TH***.

II

I will introduce, now, Kripke's fixed point construction (as in Kripke (1975)) using the supervaluational scheme. As before N will be a model of the base language with domain D, $\langle N , A\rangle$ refers to the model of the expanded language L+ whose interpretation of Tr is A, which will be a set of (codes of) formulas of L+. Again, I will use $|\alpha|M = 1$ to mean that the formula α has semantic value 1 in the model M (and the same for having semantic value 0); thus $|\ |$ is a classical valuation. *SENT* will be the set of (codes of) sentences of L+.

Let us first define the supervaluational scheme, which is a third valued valuation $|\ |s$ that will take as semantic values 1, 1/2 and 0. For any $\psi \in$ L+, any model N for the base language L and any set of (codes of) sentences X, $|\psi|s\langle N,X\rangle$ is defined in the following way:

$|\psi|s\langle N,X\rangle = 1$ iff, for every Y, such that $X \subseteq Y \subseteq SENT - X^-$, $|\psi|\langle N,Y\rangle = 1$;
$|\psi|s\langle N,X\rangle = 0$ iff, for every Y, such that $X \subseteq Y \subseteq SENT - X^-$, $|\psi|\langle N,Y\rangle = 0$;
$|\psi|s\langle N ,X\rangle = 1/2$ otherwise.

We can define now the following series of sentences of L+, for any ordinal

$VF0 = \emptyset$
$VF\sigma+1 = \{\varphi \in L+ : |\varphi|s\langle N,VF\sigma\rangle = 1\}$
$VF\lambda = \cup_{i<\lambda} VFi$

where λ is a limit ordinal.

As in the case of the previous section, we need to show, first, that the construction in monotonic.

Lemma 2.1 (Monotonicity, Kripke (1975)) *If $\vartheta \leq \rho$, then $VF\vartheta \subseteq VF\rho$.*

For the same considerations as in Theorem 1.2 there will exist a fixed point of the construction, that is an ordinal ρ such that $VF\rho = VF\rho+1$. I will call this fixed point, ***VF***.

Following Kripke (1975) and Field (2008) we can now define variations on the supervaluational scheme by imposing a condition Φ on the candidate extensions of the truth predicate. These restrictions will create other fixed points that will be supersets of ***VF***. In order to proceed, we define $|\psi|\Phi,s\ \langle N ,X\rangle$ more generally:

$|\psi|\Phi,s = 1$ iff, for every Y, such that $\Phi(Y)$ and $X \subseteq Y \subseteq SENT - X^-$, $|\psi|\langle N,Y\rangle = 1$;
$|\psi|\Phi,s = 0$ iff, for every Y, such that $\Phi(Y)$ and $X \subseteq Y \subseteq SENT - X^-$, $|\psi|\langle N,Y\rangle = 0$;
$|\psi|\Phi,s = 1/2$ otherwise.

In this definition I am presupposing that there will always be a Y satis- fying the condition Φ and such that $X \subseteq Y \subseteq SENT - X^-$. Given a condition Φ, I will call $VF\sigma\Phi$ the σ stage of the construction using Φ as the property to be satisfied by the candidate extensions of the Truth predicate. I will call ***VF***Φ the fixed point of such construction.

We can consider now the following fixed points corresponding to the following conditions:

The vacuous condition: ***VF***
Consistency: ***VF****c*
Closure under classical deduction: ***VF****cd*
Maximal consistency: ***VF****mc*

A trivial generalization of Lemma 2.1 together with the considerations in Theorem 1.2 show that all of ***VF***, ***VF****c*, ***VF****cd* and ***VF****mc* exist. We must see now that all these fixed points are consistent. The following Lemma offers a sufficient condition on Φ for consistency.

Supervaluations and Horwich's Fixed-Point Theory of Truth

Lemma 2.2 (Field (2008), p. 180) *Let λ be the Liar sentence. For any given property Φ, if for every consistent and deductively closed set of sentences Z such that λ Ï Z and ¬λ Ï Z there are Y1 and Y2 such that λ Ï Yi, λ∈Y2, Φ(Yi) and Z ⊆ Yi ⊆ SENT−Z− (1 ≤ i ≤ 2), then **VFΦ** is consistent.*

Corollary 2.3 (Field (2008), p. 180)

*(i) **VF** is consistent.*
*(ii) **VF**c is consistent.*
*(iii) **VF**cd is consistent.*
*(iv) **VF**mc is consistent.*

There are several relations that can be established between the fixed points we have presented.

Proposition 2.4
*(i) **VF** **VF**c (Kripke (1975, page 711))*
*(ii) **VF**c **VF**dc*
*(iii) **V F**dc **V F**mc (Kripke (1975, page 711))*

Finally, we can now see that ***H*** is consistent, as it is just ***VF***.

Lemma 2.5 *For any ordinal σ, Hσ ⊆ VFσ+1.*

Lemma 2.6 *For any ordinal σ, VFσ ⊆ Hσ.*

Corollary 2.7 ***H** = **V F***

We can see now that interpreting the truth predicate as any extension of ***H*** such that is disjoint with ***H***− provides us with a model for the fixed point.

Lemma 2.8 *For all sets of sentences of* L+*, Y, such that **H** ⊆ Y ⊆ D− **H**−, |**H**|⟨N,Y⟩ = 1.*

Corollary 2.9 *For all sets of sentences of* L+*, Y, such that **H** ⊆ Y ⊆ D−**H**−, |T**H**|⟨N,Y⟩ =1.*

We can also prove an important feature of Horwich's fixed point.

Proposition 2.10 *For all sentences φ of* L+*, φ ∈ **H** if, and only if, Tr φ ∈ **H**.*

Note also that, as the following Proposition shows, all the axioms of the Minimalist theory are founded, that is they are in ***H***.

Proposition 2.11 *T**H** ⊆ **H** .*

III

We should ask ourselves now whether the theory of truth Horwich is propos- ing is satisfactory. Recall that we can now characterize in a precise way which is, according to Horwich (2001), the minimalist theory of truth: T***H***. This means, according to Horwich, that all that can be said about truth should follow from T***H*** ∪ *E*, where *E* should be something on the lines of a theory of all non-truth facts. Hence, we could take the set T = {*φ* : T***H*** ⊢ *φ*} to be everything that can be said with respect to pure truth in Horwich's picture.

It is well known that ***VF*** has many unsatisfactory properties, which, as we will see in the following lines, are inherited by T. Here there are four laws we might expect to have in T:

1. For any sentence *x*; *Tr* ⟨¬¬*x*⟩ if, and only if, *Tr* ⟨*x*⟩ .
2. For any sentences *x*, *y*; *Tr* ⟨*x* ∨ *y*⟩ if, and only if, *Tr* ⟨*x*⟩ or *Tr* ⟨*y*⟩.
3. For any sentences *x*, *y*; *Tr* ⟨*x* ∧ *y*⟩ if, and only if, *Tr* ⟨*x*⟩ and *Tr* ⟨*y*⟩.
4. For any sentences *x*, *y*; if *Tr* ⟨*x* → *y*⟩ and *Tr* ⟨*x*⟩ , then *Tr* ⟨*y*⟩.

⟨N , *A*⟩ Unfortunately, though, none of these laws are satisfied by T; specifically, the problem is that we can find some instances that are not in it.

We can see now some of the difficulties that the Liar poses to the minimalist theory of truth with all its virulence. First, all the principles 1-4 will not be in T, which means, according to Horwich, that they will be principles about truth that will remain unknown to us; as a matter of fact, they will be conceptually impossible to know. Moreover, since 1-4 are not in ***H***, their instances of the T-schema are not in the minimalist theory of truth (that is, they are not in T***H***) and, hence, even if they were in T, they could not be declared true.

Let us see what does exactly mean to say that laws 1-4 are not in ***H***. We have seen that any model M for L+ with an extension of the truth predicate, *Tr*M, is such that ***H*** ⊆ *Tr*M ⊆ *D* − ***H***− satisfies T***H*** –which is the minimalist theory of truth– and ***H***. We have to ask ourselves, first, who should we understand ***H*** is. The way the construction is devised makes it natural to consider the set ***H*** ∪ ***H***− as the set of grounded sentences; specifically, ***H*** is the set of determinately true sentences (that is, supposing there are not vague predicates nor other sources of indeterminacy in Horwich's epistemic sense, the sentences which are conceptually possible to know) and ***H***− is the set of determinately false sentences (that is, the sentences whose negations are determinately true). Recall that all of these are relative to a given ground model N . To continue with this picture, note that, as I said, we can interpret *H*0 as the theory of all non-truth facts given by the ground model N and define ***T*** = {*φ* : T***H*** ∪ *H*0 ⊢ *φ*} as everything that can be known about truth at all. Then, it is natural to expect the following proposition (for any given sets of sentences Γ and Δ, I will use Γ |= Δ to mean that, for every *δ* ∈ Δ, Γ |= *δ*).

Supervaluations and Horwich's Fixed-Point Theory of Truth

Proposition 3.1 ***T*** **=** ***H***

Thus, ***H*** contains everything we can know about truth at all. So the fact that principles like 1-4 are not in ***H*** is a major problem for minimalism.

Can this situation be ameliorated? Yes, it can. It is well known that other fixed points of the supervaluational scheme, created using restrictions on the candidate extensions of the truth predicate, are much better behaved with respect to principles like 1-4. So the question is whether Horwich's construction can be manipulated so that the fixed point we get at the end be stronger than ***H***; ideally the one constructed imposing maximal consistency to the candidate truth extensions, call it ***VF****mc*. This manipulation, though, should be made following independent reasons beyond the fact that ***H*** does not contain principles 1-4. This can be done, if we have in mind Horwich's position in front of the Liar paradox.

We can ask ourselves, now, which model for the expanded language is the *actual* model; which model captures the actual world. First, since Horwich's position in front of the Liar defends, that any sentence (in particular the Liar) is such that it is true or it is false (that is, it is not true), the extension of the truth predicate in the actual model will have to be, at least, complete; there will not be undecided cases of being true.

On the other hand, Horwich adopts classical logic, which means that not only all sentences are either true or false, but also that it not the case that they are true and false. Hence, it seems natural to expect from the extension of the truth predicate to be consistent; that is, no sentence is both true and false. All this means that the extension of the truth predicate in the actual model should be, at least, maximally consistent. On the other hand, by factivity of knowledge, it is reasonable to expect the extension of the truth predicate in the actual model to be a maximally consistent superset of ***H***.

These considerations naturally suggest to restrict our attention, in gen- eral, to models whose extension of the truth predicate is maximally consis- tent. Hence, we can bring this restriction to the consequence relation used in the construction.

Following this line of thought I will call a model M for L+ mc-acceptable, in symbols M*mc* if, for every $\varphi \in$ L+, either φ or $\neg\varphi$ belong to *Tr*M*mc* but not both. We can restrict, then, logical consequence to mc-acceptable models; that is, given a set of sentences Γ and a sentence α of L+, $\Gamma \models_{mc} \alpha$ if, and only if, for every mc-acceptable model M*mc*, if $|\Gamma|M_{mc} = 1$ then $|\alpha|M_{mc} = 1$.

The definition of the new series will be the same but substituting the unrestricted consequence relation by $\models_{mc}$. Lemma 1.1 can be proved in the same way (essentially it uses the fact that, if $\varphi \in A$, for any sentence φ and set of sentences *A*, then $A \models_{mc} \varphi$) and, hence, a fixed point of the construction, let us call it ***H****mc*, will exist. Lemmas 2.5 and 2.6 can easily be proved for ***H****mc* and ***VF****mc* to show that ***H****mc* = ***VF****mc*.

Now all the instances of laws 1-4 are validated in ***H****mc* and we have found some independent reasons –that is, Horwich's stance in front of the Liar– to motivate the adoption of this stronger fixed point.

Not everything are good news, though. Recall that some of the instances of the T-schema are not in *T****H***, which means that the utility of the truth predicate is seriously impaired. Moreover, this can be known from inside the model. For note that, given $\lambda \leftrightarrow \neg Tr\langle\lambda\rangle$, the $Tr\langle\lambda\rangle$ -instance of the LEM, that is, $Tr\langle\lambda\rangle \vee \neg Tr\langle\lambda\rangle$, is equivalent to $\neg(Tr\langle\lambda\rangle \leftrightarrow \lambda)$. Now, since both $\lambda \leftrightarrow \neg Tr\langle\lambda\rangle$ and $Tr\langle\lambda\rangle \vee \neg Tr\langle\lambda\rangle$ are theorems, we have that $\neg(Tr\langle\lambda\rangle \leftrightarrow \lambda)$ is also a theorem and, hence, it is in ***H*** and in T. Hence, it is not only conceptually possible to know that the Liar-instance of the T-schema is just false, but this is a fact about pure truth.

References

Field, Hartry (2008). *Saving Truth From Paradox*. Oxford: Oxford University Press.

Horwich, Paul (1998). *Truth*. 2nd. Oxford: Oxford University Press.

—(2001). "A Defense of Minimalism". In: *Synthese* 126.1-2, pages 149–65. Reprinted in Paul Horwich (2010b). *Truth-Meaning-Reality*. Oxford: Ox- ford University Press, pages 35–56.

—(2010a). "A Minimalist Critique of Tarski". In: *Truth-Meaning-Reality*. Oxford: Oxford University Press, pages 79–97.

—(2010b). *Truth-Meaning-Reality*. Oxford: Oxford University Press.

Kripke, Saul A. (1975). "Outline of a Theory of Truth". In: *Journal of Philosophy* 72.19, pages 690–716.

III CONFERENCIA DE GRADUADOS DE LA SLMFCE

Symbiosis research and natural selection [1]

Javier Suárez [2]

Departing from the idea that any entity that exhibits heritability, variation and fitness is a candidate for being naturally selected, abstract models of natural selection are normally classified in two different groups: On the one hand, what we could call the replicator/interactor framework developed by Dawkins and Hull; on the other hand, the so called "received view", originally introduced by Lewontin and recently developed by Peter Godfrey-Smith. In this paper, I will argue that both approaches fail to capture the selective nature of holobionts, i.e. they are not able to explain why symbiotic organisms are units of selection and can be naturally selected. Hence, I will show that both the replicator/interactor framework and Godfrey-Smith's theory are insufficient for totally capturing the abstract nature of natural selection in its full range of application. Finally, I will suggest that the new models should be less centred in the idea of reproduction and more centred in the ideas of persistence and self-maintenance.

Keywords: units of selection – symbiosis – Hologenome Theory of Evolution – holobiont – interactor/replicator – Darwinian individual

Introduction

During the last decade, Philosophy of biology has been experiencing what could be referred to as an internal revolution. Due to the recent interest that philosophers have shown in phenomena such as niche construction theory, eco-evo-devo, systems biology or research on the phenomena of symbiosis, traditional categories (biological individuality, inheritance, fitness, and so on) have been proved to be very limited and to need a rethink. Particularly interesting among these phenomena is the recognition of the pervasiveness of symbiotic associations among living beings, which has led philosophers to an important shift in many of their preconceived ideas on biological individuality, the notions of reproduction and inheritance, the definition of fitness or the tree of live hypothesis, among others. The purpose of this paper is precisely to understand another possible dimension that symbiosis research has, namely: its role as a hallmark of the (in–) adequacy of certain models of natural selection. More precisely, in this paper I will delimit a subcategory among the biological associations traditionally referred as "symbiotic" –holobionts– and I will argue that they pose several issues for the two dominant theories of natural selection: First, for the replicator/interactor model and second, for what could be called "received view" on natural selection, as has recently been developed by Peter Godfrey-Smith. [3] [4]

The structure of the paper will be as follows: First, I will introduce the idea of symbiosis as applied to holobionts and I will argue that holobionts fulfil all the basic and more abstract requirements for being considered units of selection, namely: they exhibit heritability, fitness and traits that are subject to phenotypic variation; second, I will argue that the replicator/interactor framework is not a suitable framework to capture the selective nature of holobionts, since the two notions collapse; third, I will address Godfrey-Smith's notion of Darwinian individual and I will contend that it is not able to capture the selective nature of holobionts, since the requirements he demands for an entity to be considered as a unit of selection are very restrictive; finally, I will suggests that we need new models of natural selection less centred in the idea of reproduction and more centred in the ideas of persistence and self-maintenance.

1. Symbiosis in the context of the Hologenome Theory of Evolution

The importance of the notion of symbiosis and symbiosis research acquires a particularly significant dimension in the context of the so called Hologenome Theory of Evolution (in advance, HTE) (Zilber-Rosenberg & Rosenberg 2008; Rosenberg & Zilber-Rosenberg 2013). The main tenet of that theory is to show how the traditional conception of biological individuality has been misguided due to the underappreciation of the extant interactions among superior metazoans and plants and the symbiotic microorganisms they interact with or, in other words, as a consequence of the underappreciation of the intimate biological interactions extant among different organisms[5]. Minimally, these biological interactions would involve two organisms of different species: a host and a symbiont –though they can also involve more than one symbiont– that interact in a concrete intimate manner.

The most important claim of the HTE is the assertion that "each holobiont (host + microbiota), with its hologenome (host genes + microbiome), is a unique biological entity, with the sum of the dynamic interactions within the holobiont giving rise to the genotype and phenotype of the organism, as we know it. The hologenome concept posits that the holobiont (host + all associated microorganisms, including viruses), being a unique biological entity, acts also as a level of selection in evolution" (Rosenberg & Zilber-Rosenberg

2013: viii). To say that holobionts act as units of selection in evolution entails the acceptance of the fact that they satisfy the set of necessary and sufficient requirements for an entity to be susceptible to evolve by natural selection. In other words, this means to accept that: (1) holobionts exhibit specific holobiont-level adaptations (i.e. they are susceptible of variation), (2) holobionts bear emergent fitness, irreducible to the fitness of the interacting organisms and (3) their characteristics are heritable or, in other words, holobionts give rise to new holobionts.[6]

The new context opened by HTE demands for a new definition of symbiosis that would encompass only the cases of holobionts. I have proposed such a definition elsewhere (Suárez, unpublished). Following such definition, we can say that a symbiotic association is any kind of biological interaction among two organisms of different species (parasitic, commensalist or mutualistic) which, first, satisfies certain topological and temporal conditions and, second, exhibits new emergent traits or adaptations with respect to the interacting organisms.[7] On the one hand, the topological and temporal conditions demand the relationship to be *intimate* –i.e. the organisms must be in close physical contact with each other– and *constant* –i.e. the union has to last "at least a substantial proportion of the lifespan of the interacting organisms" (Douglas 2010: 9). On the other hand, the exhibition of emergent traits, i.e. traits given at the specific level of the holobiont, would prove that the interacting organisms exhibit a common evolutionary fate and a high degree of functional organization.

This definition of symbiosis is essential to show the evolutionary role that holobionts can play, since it shows the requirements that they fulfil and that ask for their consideration as units of selection.

In the next section, I will show how the new definition of symbiosis in the context of the HTE and the existence of holobionts poses problems for the extant accounts of natural selection.

2. The replicator/interactor framework as a deficient account of holobionts

Richard Dawkins (1976) and David Hull (1980) famously characterized the process of evolution by natural selection as the consequence of the differential survival of two distinct entities: replicators and interactors. According to Hull (1980: 318), a replicator is "an entity that passes on its structure directly on replication", whereas an interactor is "an entity that directly interacts as a cohesive whole with its environment in such a way that replication is differential". Natural selection would thus be "the process in which the differential extinction and proliferation of interactors cause the differential perpetuation of the replicators that produced them" (Hull 1980: 318).

It is important to take into account the assumptions and implications of the view, in order to fully understand its scope. To start with, it assumes that replicators do not need to experience any process of reconstruction when they are passed on to the next generation. The canonical example here would be genes that, though they create a copy of themselves in order to form the double helix, there is always one chain that passes on intact to the next generation. A corollary of this requirement is temporal stability, as Dawkins (1976) famously claimed: replicators need to be temporally stable, they have to persist through time so as to natural selection can act on them. Second, it assumes that processes such as meiotic drift, crossing over, etc. that bias the ways in which replicators are passed on –since they are not passed exactly intact– do not occur. Finally, it assumes a simple picture of development in which interactors –or vehicles– would exclusively appear as a consequence of the "constructive power" of replicators. Once created, those interactors would be the entities that turn out to relate with the environment, and depending on their success in doing so, replicators would or would not pass on to the next generation –i.e. they would or would not proliferate.

My argument against the replicator/interactor framework is based on the well documented case of aphids and their obligate endosymbionts, bacteria *Buchnera aphidicola* (Moran 2006; Dale & Moran 2006). The main features of this symbiotic association are: first, the fact that it is a case of endosymbiosis –i.e. the bacteria reside inside a particular organ of the aphid, called bacteryocite; second, the fact that it is obligate for the two organisms– there are no *Buchnera*-free aphids, nor aphid-free Buchnera either; third, the fact that *Buchneras* are vertically transmitted, i.e. directly from the mother to the offspring (Bright & Bulgheresi 2014, for a review of the mechanisms of transmission).

The replicator/interactor framework has many problems to deal with the aphid-*Buchnera* consortium. First, the notion of replicator seems to be useless for the case of the *Buchnera* since, on the one hand, when the holobiont reproduces *Buchnera* pass on their structures intact, i.e. the new holobiont acquires entire *Buchneras* from its progenitor; but, on the other hand, *Buchneras*, in contrast with replicators –and violating the definition aforementioned– need to be rebuilt in every new cycle, since they are bacteria. So, although from the perspective of the holobiont they seem to be replicators, *Buchneras* violate the basic criterion in

the definition of replicators and also lack temporal stability.

Second, and in connection with the last criticism, it seems very implausible that we could find any candidate for the role of replicators in the case of the holobiont. Since holobionts violate the traditional and simple picture of development that the replicator/interactor framework presupposes, the identification of replicators seems very implausible: are the genes of the aphid replicators? Then we would lack something that is necessary: *Buchneras*. Are also *Buchneras* replicators? If so, then we are trapped back in the first problem.

Third, *Buchneras* can difficultly be considered as interactors, since they do not really relate with their environments: the entity that interacts with the environment causing reproduction to be differential is the holobiont. But, again, if the consortium is the interactor… where are replicators?

For these reasons, united to the fact that instances like the consortium aphid-*Buchnera* are abundant, it seems to me that the replicator/interactor framework is not useful as a universal model of natural selection. In the next section I will argue that Godfrey-Smith's theory is not better suited to address the problems posed by holobionts.

3. The notion of "Darwinian individual": holobionts are not reproducers

Godfrey-Smith's approach to natural selection departs from the basic abstract scheme according to which "evolution by natural selection is the large category of change due to variation, heredity and reproductive differences" (Godfrey-Smith 2009: 39). More precisely, he defends that the kind of entities that satisfy those very abstract requirements are *Darwinian populations*, i.e. the "collection of causally connected individual things" that he refers to as *Darwinian individuals* (2009: 39). Darwinian individuals, Godfrey-Smith claims, are entities that have the capacity and ability of self-replication; in other words, Darwinian individuals are *reproducers*. According to Godfrey-Smith, reproducers can be of three different types, namely: simple, collective or scaffolded:

> A *simple* reproducer is something that can give rise to more objects of the same kind largely through the operation of resources internal to it –through its own biological machinery, in a broad sense– and, further, is not made of smaller parts that also have this capacity. (…) A *collective* reproducer is a reproducing object that has parts that are themselves simple or collective reproducers. (…) Third, a *scaffolded* reproducer is an entity that reproduces (or is reproduced) in a way highly dependent on resources external to itself. (Godfrey-Smith 2015: 10121)

Analogously to the replicator/interactor framework, Godfrey-Smith's theory cannot justify why certain holobionts are units of selection, since the criteria he offers for being a reproducer are too restrictive. In this case, I will analyse the example of a developmentally induced symbiosis: mice and *Bacteroides thetaiotaomicrom*. In a very famous experiment, Stappenbeck et al. (2002) showed first, that the acquisition of *B. thetaiotaomicrom* is necessary for blood vessel formation in mice and, second, that those bacteria could be acquired *in any moment* of the developmental process–acquisition would be immediately followed by the development of the blood vessel system. Of course, mice without a normal blood vessel system cannot survive except in lab conditions, which suggest the necessity of the acquisition of the *B. thetaiotaomicrom* and thus its consideration as a holobiont. It is very important to notice that mice acquire their symbiotic bacteria directly from the environment, through a process of horizontal acquisition. The question now is, what kind of reproducers, if any, would the consortium mouse-*Bacteroides* be?

First, it seems very implausible to consider the holobiont as a simple reproducer, since it is pretty clear that the holobiont has smaller parts that can themselves replicate. Second, it does not seem plausible to consider it as a scaffolded reproducer. Paradigmatic scaffolded reproducers, like viruses, are characterized for the fact that they need to parasitize another organism in order to replicate, since they lack all the cellular machinery that is required to do so. The case of the consortium mouse-*Bacteroides* does not seem similar at all, since the holobiont does not need to parasitize any other organism or external source in order to self-replicate.

What about considering this case as a case of collective reproduction? I do not think this route is plausible either, since the reproduction of the holobiont, and also the reproduction of the mouse that is a member of the holobiont, is entirely dependent on the normal existence of the consortium mouse-*Bacteroides*. It does not seem plausible to me to say that the holobiont is a sum of parts –namely: mouse + *Bacteriodes*– that when put together they simply acquire the status of a new reproducer in a higher level of the biological hierarchy. It seems that the holobiont exhibits emergent properties that each of its members lack, and therefore it is difficult to take it as a case of collective reproduction.

For these reasons, I think that Godfrey-Smith's account is not suitable for capturing the nature of holobionts either, and thus it is not a complete model of evolution by natural selection.

Symbiosis research and natural selection

4. Conclusion

In this paper, I have casted certain doubts about two very influential accounts of natural selection on the basis of the HTE. Particularly, based on the existence of holobionts and their role as units of selection, I have argued that neither the replicator/interactor framework, nor Godfrey-Smith's theory of Darwinian individuals are proved to be useful to understand the very nature of these entities as susceptible of being naturally selected. The main argument is based on the restrictive notion of reproduction that both the replicator/interactor and Godfrey-Smith's framework take.

The situation now asks for the consideration and formulation of new abstract models of natural selection that could account for the case of symbiotic organisms and their role as units of selection. My intuition –and the lessons that seems to follow from the puzzling cases aforementioned– is that we need an account that is less centred in reproduction. As it has been argued along the paper, both approaches seem unsuitable for capturing holobionts as a consequence of the importance they give to reproduction, be it conceptualized as "replicators" or "reproducers". Maybe an approach that takes into account the ideas of self-maintenance and persistence would be more suitable for capturing the essence of evolution by natural selection. Nonetheless, developing these ideas is out of the scope of this paper.

[1] A previous version of this paper was presented in 2016 meeting of Philosophy of Biology in the UK and in the III Graduate Conference of the Spanish Society for Logic, Methodology and Philosophy of Science. I would like to thank all the participants there for their feedback and helpful comments, and especially John Dupré, Jose Díez and Cristian Saborido for their careful reading and feedback on an expanded version of this paper. Finally, "Fundación Bancaria la Caixa – Becas de Postgrado: Programa Europa" is formally acknowledged for its financial support.

[2] Egenis, The Centre for the Study of the Life Sciences, University of Exeter, Byrne House, St German's Road, Exeter, EX4 4PJ, UK. Email: jsuar3b@gmail.com.
Javier is a first year PhD student under the supervision of John Dupré (Egenis, University of Exeter) and Jose Díez (LOGOS, University of Barcelona).

[3] The distinction between those two general abstract models of natural selection relies on Okasha (2006) and Godfrey-Smith (2009), who refers to the received view as the "classical tradition" of natural selection.

[4] It is important to note that I am not raising a criticism against the received view in its more abstract formulation – indeed, I take it as something that holobionts satisfy and, thus, as evidence for considering holobionts as units of selection. My criticism is directed towards Godfrey-Smith's application of the received view.

[5] In this paper, the expressions "biological interaction" and "biological association" will be used as synonymous to refer to whatever union between organisms of different species. This notion will contrast with my definition of "symbiosis", which would encompass a subclass among these associations.

[6] These are the minimal requirements for an entity to be a unit of selection, namely: the entity has to exhibit heritable variation in fitness. For a review of the minimal conditions see (Lewontin 1970), (Sober 2000) and (Brandon 2014).

[7] My definition of symbiosis is similar to –and inspired by– the definition found in (Zook 2015: 48). I won't discuss here the similarities and differences here, since it is above the scope of this paper.

References

Brandon R (2014) Natural Selection. In *The Stanford Encyclopaedia of Philosophy*, ed. EN Zalta.

Bright M & S Bulgheresi (2014) A complex journey: transmission of microbial symbionts. *Nat Rev Microbiol* **8**: 218-230.

Douglas AE (2010) *The symbiotic habit*. Princenton University Press.

Godfrey-Smith P (2009) *Darwinian Populations*. Oxford University Press.

Godfrey-Smith P (2015) Reproduction, symbiosis and the eukaryotic cell. *PNAS* **112** (33): 10120-10125.

Hull D (1980) Individuality and selection. *Annual Review of Ecology, Evolution and Systematics* **11**: 311-32.

Lewontin RC (1970) The units of selection. *Annual Review of Ecology, Evolution and Systematics* **1**: 1-18.

Moran N (2006) Symbiosis. *Current Biology* **16** (20): R866-R871.

Moran N, JP McCutcheon & A Nakabachi (2007) Genomics and evolution of heritable bacterial symbionts. *Annual Review of Genetics* **42**: 165-190.

Symbiosis research and natural selection

Okasha S (2006) *Evolution and the Levels of Selection*. Oxford University Press.

Rosenberg E & I Zilber-Rosenberg (2013) *The Hologenome Concept*. Springer.

Sober E (2000) *Philosophy of Biology*. Westview Press.

Stappenbeck TS, LV Hooper & JI Gordon (2002) Developmental regulation of intestinal angiogenesis by indigenous microbes via Paneth cells. *PNAS USA* **99**: 15451-15455.

Suárez, J. (unpublished): Mutualism, symbiosis, and symbiogenesis. Contribution presented at the *VI Congress of Philosophy of Biology and Cognitive Science*. Available online: https://pbcs6barcelona.files.wordpress.com/2015/10/abstract_jsuarez_pbcs6_long.pdf.

Zilber-Rosenberg I & E Rosenberg (2008) Role of microorganisms in the evolution of animals and plants: the hologenome theory of evolution. *FEMS Microbiology Review* **32**: 723-735.

III CONFERENCIA DE GRADUADOS DE LA SLMFCE

Natural Selection and Complexity

Giorgio Airoldi[1]

Abstract
The alleged Darwinian fundamental tenet that Natural Selection explains the complexity of life is controversial. While adaptationist narratives aims at formalizing it based on the hypothesis that phenotypic traits result from fitness optimization, mechanisms alternative to Natural Selection have been proposed to explain the tendency of organisms towards increasing complex designs. The thesis of this paper is that, in order to clarify to which extent Natural Selection or other forces are the relevant *explicans*, it is necessary to recognize that there is a wide range of radically different evolutionary facts, to which no univocal definition of complexity increase is applicable. To classify them, we introduce a bi-dimensional space by adding robustness to fitness as the second dimension to measure design and design changes. In this space, it is possible to distinguish the impact of selective and non-selective mechanisms behind each class of evolutionary fact, and the correspondent different classes of complexity increase.

Key words: Natural Selection, design, fitness, robustness, complexity, adaptationism

Darwin thought that the mechanism of Natural Selection, improving the fit between the individual and the environment, explains at the same time the variety and the complexity of living organisms (chapter 3 and 4 of the 1872 edition of *The Origin*). The first of these claims is widely accepted and formalised by Population Genetics, which explains the evolution of phenotypic variety through changes of alleles' frequencies in a population. Acceptance of the second is not so universal (see e.g. Gould & Lewontin 1979). This is partly due to its lack of formalization: its advocates usually recur to arguments that explain phenotypic traits through narratives presupposing the action of some optimization mechanism. Due to their informal nature, these arguments are impossible to falsify.

Additionally, the polysemy of the concept of complexity makes it difficult to identify what exactly needs explaining. The Science of Complexity approach, for example, identifies complexity as an emergent phenomenon that rises from the interaction among homogeneous and relatively simple elements that constitute open, non-hierarchical and far-from-equilibrium systems, in the edge between order and disorder (Mitchell 2009, Johnson 2010). The HOT (*Highly Optimized Tolerance*) approach, on the contrary, characterizes complex systems as composed by heterogeneous elements organized in hierarchical structures, that constitute organizations robust against expected turbulences, yet fragile against unexpected ones (Carlson & Doyle 2002). On the opposite side, there are 'minimalist' definitions of complexity. McShea and Brandon (2010) reduce it to the number of parts of an organism, regardless of any consideration around their origin and function. Grafen (2007, 2014), without providing a formal definition, seems to identify complexity as the remote cause of increments in phenotypic fitness, whose immediate cause is Natural Selection.

This double difficulty (to define complexity and to give a formal adaptationist explication of the tendency of organism towards more complex designs) has given rise to two lines of research, each proposing different solutions to these problems. The following table summarises their main claims.

Line of research	Answer to problem 1: Does Natural Selection formally explain Complexity	Answer to problem 2: What is complexity?	Proposals
1. Adaptationism	Main *explicands*	Simple, partial definitions (value of a trait or mix of trait, e.g. fitness)	Optimization Programs Formal Darwinism Project
2. Alternative Explications	Minor role, other more relevant mechanisms	Comprehensive, holistic definitions	Genetic Mechanisms Phenotypic Mechanisms Systemic Mechanisms

The first line of research tries to translate into formal models the adaptationist narratives, considering only the action of Natural Selection and relying on a limited concept of complexity. *Optimization programs*, for example, that represent an interesting approach borrowed from economics (Parker & Maynard Smith 1990), reduce complexity to the value of a phenotypic variable. In the optimal foraging models, where such programs have been successfully applied, this variable measures the average foraging time in a place before the individual moves to a new one, and its value is deduced from considerations around the maximization of energy assumption per unit of time (Charnov 1976). As a general trend, the narrative that explains a trait is formalized based on the assumption of the maximization of fitness (to which the trait contributes), assumption that Population Genetics usually denies. Grafen's *Formal Darwinism Project* (Grafen 2007, 2014) aims at solving this conflict and represents the most ambitious among adaptationist formalization attempts. Applying an optimization approach to population genetics equations, it shows that, at equilibrium, genetic frequencies as forecasted by these equations lead, as a general tendency, to the maximization of fitness (even if such maximization is not reached because of genetic constraints).

An extensive literature denies the assumptions of adaptionist models that Natural Selection can explain all traits (Maturana & Varela 1980, Pigliucci 2008) and that it shows unlimited capacity to produce new traits (Wagner 2015, Moczek 2008, Eldredge & Gould 1972). Based on these critics, the second line of research rejects explanations based purely on Natural

Natural Selection and Complexity

Selection and, instead of striving to formalize the Darwinian argument, looks for alternative mechanisms, non-lineal and non-progressive, to explain the appearance of complex new traits and novel architectures. We classify these proposals in three groups, depending upon where the source of new traits is identified: in genetic, phenotypic or systemic mechanisms.

Wright's *shifting-balance theory* (Wright 1982) suggests that new traits arise, in relatively short evolutionary time spans, from genetic drift in population of small dimension and geographically isolated, through the casual fixation and loss of alleles. Given that the relationship between alleles and phenotypic traits is complex and non-additive, novel phenotypes can arise even without the contribution of mutations. The mechanism of 'mutation plus Natural Selection' cannot explain the most significant phenotypic novelties because it needs too long timescales, because mutations are usually lethal and because, being a self-finalising process, it tends to destroy variety. The polymorphism of many species is the final equilibrium point of a selective process, and not the source for future adaptations. The *Punctuated Equilibria theory* (Eldredge & Gould 1972) appeals to the similar idea that phenotypic changes are rapid and limited to marginal groups, and that, if successful, they later spread to the rest of the population. This pattern of allopatric speciation, followed by geographical expansion, explains gaps in the fossil register. Wagner (2015) identifies *genotypic networks* within which potential genotypes, although differing from one another for some elements (e.g. proteins that differ in one amino acid), share the same primary functions: thanks to the accumulation of cryptic mutations, the genotype can explore this network and acquire new functions without losing the original ones. This explains the sudden appearance of novel traits.

This first group of proposals, even if it distances itself from the classic hypothesis of Population Genetics (i.e. big populations, role of mutations and Natural Selection, etc.), still considers the genotype as the source of phenotypic variability. The second group abandons this gen-centred view and looks for variability at the phenotypic level. Gould and Vrba (1982) introduce the concept of *exaptation* to identify traits that acquire adaptive value without going through a process of Natural Selection. An exaptation is defined as a trait appeared as an adaptation for some original function that is later leveraged to perform a new one (e.g. the case of the original thermoregulation function of feathers, only later exploited for flight). An exaptation can also be a trait due to environmental or architectonical constraints and lacking any function, until a change in the environment grants it one (e.g. the shape of sponges and corals due to marine currents, Gould & Lewontin 1979). In both cases, the mechanism of exaptation represents a phenotypic and not a genetic source of new traits. Moreover, the more complex the organism is, the more frequent potential exaptations are (as there are more traits and as they are interconnected in more ways than in simpler organisms): this virtuous circle explains why complexity generates more complexity.Finally, the source of traits that fuels Natural Selection has been identified in processes intrinsic to the organisms. Complex systems theory considers that these are regulated by *self-organizing laws*: organisms are a particular case of such systems, and their increasing complexity a particular output of these universal laws. Complex systems tend to move towards stable states, defined 'attractors', such that, if a perturbation moves the system away from it, this tends to move back to the original attractor or to position itself in a new one (Kitano 2004, Mitchell 2009, Kaufmann 2000). Attractors shows a certain degree of intrinsic complexity, and the movement from one to another results in a global 'jump' in complexity, not reducible to the sum of many partial increases. Another group of proposals, linked to Evo-Devo research, focuses on the path from one equilibrium state to another, more than on the equilibria themselves. Such paths are defined by *development constraints* and predetermine which organisms are possible and which are not: the change to a new phenotypic architecture is therefore not gradual, slow and driven only by the environment, but relatively quick and along a limited set of evolutionary paths (Alberch 1991). McShea and Brandon (2010) postulate that a tendency towards increasing complexity (called *Zero Force Evolutionary Law*, or ZFEL) underlies all phenomena of reproduction with variation, and that the fact needing an explanation is not evolutionary change, but stasis.

In the present paper, we defend in the first place the need for a classification of evolutionary facts as a preliminary step to the analysis of the mechanisms driving them. Adaptation of the colour of the *B. betularia*'s wings in response to environmental changes is a radically different fact than the speciation of the Galapagos finches or the appearance of a novel function like flight. Not all types of evolutionary change entail an increase in complexity of the same kind and degree[2]. The following table presents some examples of such changes and the main elements that distinguish them (in terms of traits, variants of traits and functions).

fact	trait	variant	function	Design change
Change in wings' colour in *B. betularia*	Existing (wings' colour)	Existing (black)	Same (mimicry)	New mix of traits
Speciation of finches	Existing (beak)	New (longer)	Same/new (appropriateness for different types of food)	New variants
Exaptation of feathers for flying	novel	novel	Novel (flight)	New architecture

To build any classification, it is first necessary to identify its dimensions. Adaptationist models measure design as the value of fitness (for example, in Grafen 2007: 'Adaptation is design, and maximizing fitness is what organisms are designed for'). This is equivalent to considering the organism as a black box, which transforms alleles' frequencies into average fitness, omitting considerations around the internal architec-

Natural Selection and Complexity

ture of organisms. Whenever Natural Selection is the main force behind the change, a scalar variable like fitness can be a satisfactory proxy of the salient characters of design, considered as the mix of current traits under a given and fixed architecture. However, when the evolutionary fact entails new traits and a new architecture, it is necessary to open the black box and to understand how mechanisms other than Natural Selection act within it and influence the appearance of evolutionary novelties. To do this, we propose to add robustness to fitness as second dimension of the design space.

There are many definitions of robustness in the literature (for example Kitano 2004, Carlson & Doyle 2002, Pigliucci 2008, Moczek 2008). Here, we use the definition by Wagner (ch. 8 in Wagner 2011), who identifies it at the same time as (a) the ability to survive changes in the current environment and (b) the disposition to develop new traits, functions and architectures to adapt to new environments.

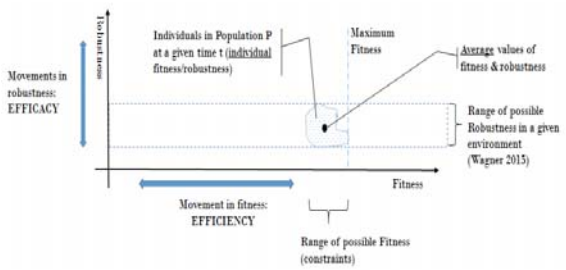


Fig. 1 – Design Space

Figure 1 shows the main elements of this design space. Each point represents an individual of the population at a given time. The set of all individuals has an average fitness and robustness, as well as a maximum and minimum of fitness (linked to genetic and other kinds of constraints) and of robustness (an individual with too low a robustness does not survive, while one with too high a robustness spends energy in traits useless in that particular environment: both are eliminated by Natural Selection). Movements along the horizontal axis towards higher fitness are linked to an *efficiency* increase: the organism fulfils the same functions, with the same traits and the same architecture, but in a more efficient way. Movements along the vertical axis towards higher robustness are linked to an *efficacy* increase: the organism realizes new functions thanks to new traits or a new architecture. Natural Selection and the other proposed mechanisms have different effects on the average and the variance of fitness and robustness, as summarised in the following table.

Force	F average	F variance	R average	R variance	Directional: Always increasing average F/R?	Continuous vs Discrete increase of average F/R ?
Natural Selection	↑	↓	0	0	yes, F	Continuous
Drift	↑↓	0	↑↓	0	no	Continuous
ZFEL	0	↑	0	↑	no	N/A (*)
Exaptations	N/A	N/A	↑	N/A	yes, R (**)	Discrete
Self-organizing rules (e.g. Kauffman's laws)	N/A	N/A	↑	↑	yes, R	Discrete

Legend
0: no impact
N/A: casual impact that is not the primary source of change

(*) averages do not increase
(**) or it would not be an exaptation

Natural Selection (according to Fisher's theorem) tends to increase average fitness and to reduce the fitness variance (Fisher 1930, Price 1973): it is a directional and continuous force. Drift acts upon averages but not upon variances: it is not directional and it is thus impossible to forecast in which direction it pushes fitness and robustness (Brandon 2006)[3]. The Zero Force Evolutionary mechanisms, on the other hand, acts upon variances but do not change averages. It is therefore likewise non-directional (McShea & Brandon 2010). The only predictable effect of exaptation (for the definition itself of exaptation as contribution to the development of a new function) is an increase of robustness (Gould & Vrba 1982). Self-organizing rules (e.g. Kauffman 2000) acts exclusively upon robustness, increasing both its average and its variance. Figure 2 shows such effects in a graphical format, detailing the distribution of the population and the correspondent average fitness and robustness before and after the action of the force.

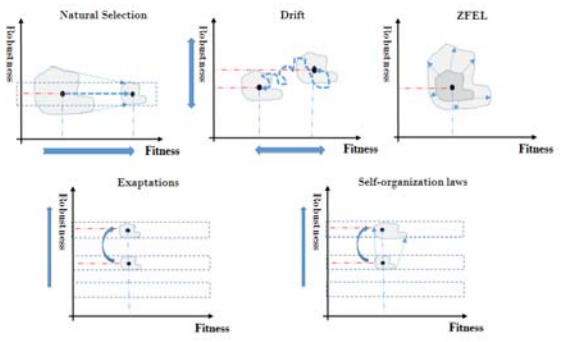


Fig. 2 – Effects on fitness and robustness of the different evolutionary forces

Natural Selection and Complexity

In the so-defined design space, we can write:

Design = f [fitness (current traits), robustness (architecture)] (1)

When an evolutionary fact is fuelled by Natural Selection alone and does not entail a change in robustness (for example, adaptation of the wing's colour of *B. betularia* to changes in the environment, or maximization of the size of a trait), fitness alone can satisfactorily capture the essence of the organism's design, as adaptationist models do. Equation (1) thus reduces to:

Design = f [fitness (current traits)] (2)

When this is not the case, and the evolutionary fact entails a change in robustness due to other non-selective forces (for example, the appearance of a new trait like feathers, or of a new function like flight), it is necessary to consider both dimensions. Figure 3 exemplifies evolutionary movements of these two kinds.

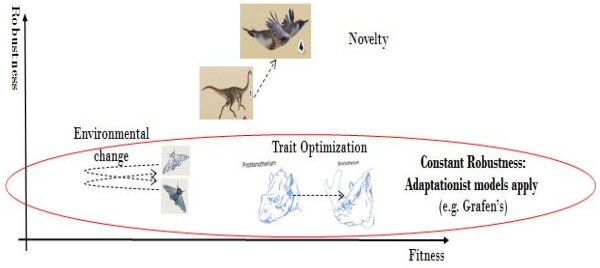


Fig. 3 – Movements in the Design Space linked to different kinds of evolutionary facts

Generic evolutionary facts can be factorized into movements along both axis. As an example, figure 4 suggests a possible explanation of the appearance of the function of flight. The first step consists in the appearance of the new trait 'feathers' caused, for example, by genetic drift. This new trait allows for better thermoregulation, thus increasing robustness, as the organism can now resist to wider ranges of external temperatures. Its impact on fitness is decided by Natural Selection, that, depending on the adaptive advantage it carries, will spread it among the individuals of the population and optimize its configuration (in term, for example, of shape and quantity of the feathers), or eliminate it. At some point, the new trait optimized by Natural Selection becomes an exaptation, allowing a primitive and rudimentary flight: robustness increases again, but not necessarily fitness. If flight does not give any competitive advantage in a particular environment, Natural Selection could push towards its disappearance (because, for example, it uselessly consumes energy resources). If flight does grant some advantage, Natural Selection tends again to its optimization (for example, in terms of shape or number of feathers, or dimensions of wings).

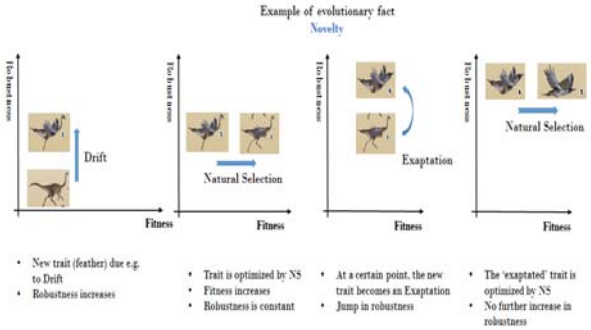


Fig. 4 – Factorizing into elementary movements of the appearance of a novelty (flight)

The movement from the initial to the final phenotype is therefore the sum of several horizontal and vertical movements, each explained by a different evolutionary force, and each entailing a change of complexity in a different sense. When robustness increases (new trait or exaptation of existing trait), the change in complexity of design entails an architectural change. When robustness remains constant, the architecture does not change either, and the evolution of design consists in the optimization of a single trait or of the mix of existing traits.

Conclusions

The Darwinian claim that Natural Selection explains the increasing complexity of organism throughout the history of life presents two problems: the lack of a universal formalization and the polysemy of the concept of complexity. While adaptationist programs try to supply such formalization reducing complexity to a phenotypic variable (usually fitness), and recurring to optimization considerations, other mechanisms have been proposed to explain the appearance of new traits and architectures. We suggest that both Natural Selection and alternative evolutionary forces contribute to phenotypic complexity, but along different axis: the former one acts mainly upon fitness, while the latter ones mainly upon robustness of organisms. Fitness and Robustness are the dimensions of a design space, in which it is possible to track evolutionary changes of very different kinds, as well as the forces behind them. Moreover, changes in fitness and in robustness are linked to different concepts of complexity. The question about whether Natural Selection explains complexity is, under this approach, simply poorly formulated.

Acknowledgments

I would like to thank prof. A. Grafen for the clarifications about his model he kindly provided me and prof. Diego Rasskin-Gutman for his comments during the presentation of this paper at the 'III Congreso de Graduados de la SLMFCE'. I would also like to thank two anonymous reviewers for their suggestions.

Natural Selection and Complexity

[1] Ph.D. Student, Department of Logic, History and Philosophy of Science, UNED Madrid,
✉ gairoldi@alumno.uned.es

[2] We think that such a classification is already implicit, for example, in the different types of 'evolvabilities' identified in Pigliucci (2008).

[3] The fact that Drift and other forces can reduce average fitness does not contradict Fisher's theorem, given that this applies when Natural Selection is the only force in place (Price 1973).

Bibliography

Alberch, P. (1991). From genes to phenotypes: dynamical systems and evolvability. *Genetica* 84, 5-11

Brandon, R. N. (2006). The principle of drift: Biology's first law. *Journal of Philosophy*, 103(7), 319-335. doi:10.5840/jphil2006103723

Charnov, E. L. (1976). Optimal Foraging, the Marginal Value Theorem, Theoretical Population *Biology*, Vol. 9, No. 2, April 1976

Carlson, J M., Doyle, J. (2002). Complexity and Robustness. *PNSA*, vol. 99, 2538-2545

Eldredge, N. & Gould, S.J. (1972). Punctuated Equilibria: An alternative to phyletic gradualism. In Schopf, Thomas J. M. (ed.). *Models in Paleobiology*, Freeman, Cooper and Company, San Francisco: 82-115

Fisher, R. A. (1930). *Genetical theory of natural selection*. S.l.: Oxford Clarendon Press.

Gould S.J., Lewontin S.J. (1979). The spandrels of San Marco and the Panglossian paradigm: a critique of the adaptationist program, *Proc. R. Soc. Lon*., B 205, 581-598

Gould, J.S., Vrba, E. S. (1982). Exaptation – A Missing Term in the Science of Form, *Paleobiology*, Vol. 8,No. 1: 4-15

Grafen, A. (2007). The Formal Darwinism project: a mid-term report, *J. Evol. Biol*., 1243-1254

Grafen, A. (2014). The formal Darwinism project in outline, *Biol Philos*, 29: 155-174

Johnson, N. (2010). *Simply Complexity: a Clear Guide To Complexity Theory*. Oneworld Publications

Kauffman, S., (2000). *Investigations*. Oxford University Press.

Kitano, H. (2004). Biological robustness. *Nature Reviews Genetics*, 5(11), 826-837. doi:10.1038/nrg1471

Maturana, H., & Varela, F. J. (1980). *Autopoiesis and cognition: The realization of the living.* Dordrecht: D. Reidel.

McShea, D., Brandon, R. (2010). *Biology's First Law: the tendency for diversity and complexity to increase in evolutionary systems*. The University of Chicago Press.

Mitchell, M. (2009). *Complexity. A Guided Tour.* Oxford University Press.

Moczek, A. P., (2008). On the origins of novelty in development and evolution, *BioEssays* 30:432–447

Parker, G., Maynard Smith, J. (1990). Optimality theory in evolutionary biology, *Nature*, 348: 27-33

Pigliucci, M. (2008). Is evolvability evolvable? *Nature Reviews Genetics*, 9(1), 75-82. doi:10.1038/nrg2278

Price, G. R. (1972). Fisher's 'fundamental theorem' made clear. *Annals of Human Genetics*, 36(2), 129.

Wagner, A. (2011), *The Origins of Evolutionary Innovations*, Oxford University Press

Wagner, A., (2015). *Arrival of the Fittest*, Oneworld Publications

Wright, S. (1982). The shifting balance theory and macroevolution, *Ann. Rev. Genet.* 16:1-19

III CONFERENCIA DE GRADUADOS DE LA SLMFCE

Entrenching the epistemological side of computer simulations: explanation and unification

Juan M. Durán
High Performance Computing Center Stuttgart
Universität Stuttgart
duran@hlrs.de

Abstract
I draw attention to philosophical issues underlying scientific explanation in computer simulations. To this end, I first identify the explanans and the explanandum; this is important since it is not clear what is being explained nor in virtue of what we explain. Second, I suggest the unificationist account of scientific explanation as the most suitable theoretical framework. Despite its suitability, however, I argue that the unificationist needs to be slightly reinterpreted in order to give room to computer simulations. Third, I discuss what it is to me the epistemic gain of explaining results of computer simulations. Finally, I present what it is still missing for a full-fledged account of explanation in computer simulations.

Keywords: Computer simulations - scientific explanation - unificationist account of scientific explanation

The philosophical debate on computer simulations emerged simultaneously with the irruption of computers into the scientific milieu. The early work of authors like Cohen (Cohen, 1961), and Naylor, Burdick, and Sasser (Naylor et al., 1967) set out the philosophical importance of computer simulations and their imprint in the experimental life at an early age. In the past few years, the philosophical attention on computer simulations has been revitalized in the work Humphreys (Humphreys, 2004), Winsberg (Winsberg, 2010), and most recently Morrison (Morrison, 2015), among others. The common denominator between the early philosophers and their contemporaries lies in the interest on analyzing the epistemic virtues of computer simulations. Current philosophical literature is particularly susceptible to discuss the status of computer simulations as -novel forms of- experimentation (Humphreys, 2004; Winsberg, 2010; Morrison, 2015), and as special kinds of mathematical models on a digital machine (Morgan, 2005; Frigg and Reiss, 2009). This article also advances on the epistemological analysis of computer simulations, although it does so by looking at their explanatory force. The aim is to frame computer simulation within a suitable account of explanation (i.e., the unificationist account), to show how an explanation is carried out, and finally what kind of understanding we obtain. The two main questions there were: how is it possible to explain results of a computer simulation? and, what kind of epistemic gain should we expect?

Typically, studies on explanation in computer simulations are less about the logic of explanation and more about referring to yet another epistemic activity carried out by simulations. In the last sense, philosophers either stress explanation as a way of obtaining knowledge of the world *vía* simulations (e.g., (Beisbart, 2012)), or equate it with other epistemic activities, such as prediction and measurement (e.g., (El Skaf and Imbert, 2012)). Of the meager remaining literature, it has been claimed that results of computer simulations can be explained by discovering the underlying mechanisms that represent real -world phenomena. This is the opinion of Ulrich Krohs who upholds a mechanistic explanation of a simulation of the Belousov-Zhabotinsky reaction (Krohs, 2008). For this to happen, Krohs calls for two elements to be in place. First, he takes that computer simulations are theoretical dynamic models that refer to the internal mechanisms of a real-world phenomenon, and which can be directly implemented on the digital computer (Hartmann, 1996). â€œSuch modelsâ€ , says Krohs, â€œmay be regarded as not only describing, but also as explaining, the process under consideration.â€ (Krohs, 2008, 278). In this sense, the simulation is regarded as â€œan analogue to the modeled system with respect to its dynamicsâ€ (Krohs, 2008, 283). Second, to explain is to exhibit the mechanisms that bring about the dynamics of the system modeled as described in the theoretical model (Krohs, 2008, 283-284). In this way, Krohs assimilates the mechanistic theoretical framework of explanation as the most suitable account for simulations. That is, he expects to explain *why* by explaining *how* (Bechtel, 2005, 422). I contend his account of explanation for computer simulations and offer my own approach.

The following is a summary of the presentation given on the 2 of June 2016 in the *III Congreso de Graduados de la Sociedad de Lógica, Metodología y Filosofía de la Ciencia en España*. The presentation is framed as follows. First, I identify the *explanans* and the *explanandum* for computer simulations. This is an important and non-trivial step since it is not obvious what is being explained (e.g., data broadly construed, the real-world phenomenon) nor in virtue of what we explain (e.g., theory, mathematical models, simulation models). To make this point clear, I present an example that will be used throughout the presentation. Second, as mentioned before, I suggest the unificationist account as the most suitable theoretical framework for computer simulations (Kitcher and Salmon, 1989). Despite its suitability, however, I argue that the unificationist needs to be minimally modified in order to accommodate computer simulations. Third, I discuss what it is to me the epistemic gain of explaining results of computer simulations. To my mind, such gain goes beyond the unificationist standard ideas. Finally, I present some of the philosophical limitations of my account, and how they can be overcome. Allow me to elaborate.

It is standard in the literature to take the *explanans* as consisting of well-confirmed scientific hypotheses, laws, theories

Entrenching the epistemological side of computer simulations: explanation and unification

and, as found in more recent literature, of scientific models. Computer simulations are not alien to this frame of reference, as there is no conceptual problem in conceiving them as some kind of *special* scientific model. Much of current philosophical literature takes the mathematical model -as implemented in the computer simulation- as the unit that carries the explanatory input (e.g., (Krohs, 2008) (Weirich, 2011)). I content this point, and defend the view that the explanans encompass the simulation model itself. If this last claim is correct, then several questions arise that need to be answered. What does ground the simulation model over a mathematical model -or a theory- as more suitable for the explanans? In addition, if the simulation model is indeed part of the explanans, are we risking some sort of self-reference between explanans and explanandum (i.e., the simulation model would be responsible for both, explaining and producing the results of the simulation)? Lastly, it has been claimed that computer simulations entail *epistemic opacity*, understood as our limitation of fully fathoming their results (or the process for obtaining such results). Does epistemic opacity have any bearing in the construction of the explanans? As for the *explanandum*, I take it to be the results of the computer simulation as data broadly constructed. Grounds for such an interpretation will be given, as well as some concerns addressed.

Identifying the explanans and explanandum are only part of the issue. Several others still await for an answer. How, if at all, could the simulation model be reconstructed in order to be part of the explanans? It seems that the conceptual requirements for reconstructing the explanans exceed any human capacity to fully grasp the simulation model. Another challenge is to deal with *conceivable computer simulations results*, such as simulating the gravitational constant to a random value. Could we talk of genuine explanatory input for such a computer simulation? Let it be noted that, at a this level of analysis, computer simulations bring into question the possibility of explaining data *tout court*, as it has been resisted by Woodward (Woodward, 1989). I will, however, not elaborate on this point. Instead, this presentation addresses objections to my own approach and the ways they could be overcome.

As for the second part of the presentation, there are several reasons that point to the *unificationist* as the most suitable account of scientific explanation for computer simulations. One advantage, for instance, is that the explanatory input is obtained by deriving a description of a phenomenon from a set of argument patterns. Such patterns, I argue, are reconstructed from the simulation model, just as well as the description of the results of the simulation. Moreover, derivation is a desirable feature since it suits well into the nature of computer simulations as abstract algorithmic structures. Now, despite the advantages offered by the unificationist, there are still issues that require our attention. A major concern is that computer simulations are prone to all kinds of errors, such as truncation errors, and round-off errors, and the like. Since they have a specific weight in the explanatory input, it becomes necessary to be able to account for them. Indeed, whereas the simulation model could be a good representation of the target system, the results could deviate from the 'true' value of the empirical phenomenon being simulated. Now, the problem is that such errors are rarely available for reconstruction simply because they are not necessarily known beforehand. Here is where the unificationist account seems to fall short, as it does not consider any schemata other than one having all the information readily available for the construction of the explanans. In addition, the standard notion of *understanding* as unifying a multiplicity of phenomena needs to be properly adapted to computer simulations. Conversely, a conceptualization of computer simulations as unifying systems is accepted. I briefly show how a derivation is done.

Other accounts of scientific explanation, more prominently the mechanicists (e.g., (Craver, 2006), (Salmon, 1984)), and mathematical explanation of physical phenomena (e.g., (Batterman, 2002)), seem to fail to accommodate computer simulations into their framework. The former fails because, I believe, computer simulations do not bear causality, although causal relationships could be represented by means of patterns. This means that we are unable to identify physical causal relations acting in the simulation, although we are able to represent them in a suitable manner. In the context of computer simulations, causal relations are represented by algorithmic structures, that is, descriptions of the acting causes, their attributes, and relations, all in a suitable programming language. Again, this does not entail that we are able to infer causation from such algorithmic structure (Freedman and Humphreys, 1999), nor that unanticipated causes are present. Let it be noted that Woodward's *manipulationist account* (Woodward, 2003) also treats causality in rather representative terms. However suitable this might seem, there are good reasons for disregarding the manipulative account as a suitable explanatory framework for computer simulations. Most of my reasons are grounded on shortcomings inherent to the manipulative account that are unsuccessful for computer simulations (see (Durán, 2013)).

The third part of this presentation addresses the epistemic gain of explaining results of a computer simulation. The unificationist takes that explaining is an epistemic enterprise *par excellence*, one that "makes the world a more transparent place" by reducing the multiplicity of phenomena that we have to take as ultimate (or brute) using the same patterns of derivation again and again. I endorse this viewpoint. Moreover, to my mind, there are no conflicts adopting this viewpoint for computer simulations. However, I also believe that explaining facilitates other epistemic activities as well, such as grasping technical difficulties behind coding more complex and realistic simulations, providing clues for verification and validation methods, among others. The presentation discusses this 'new' epistemic gain of explaining results of a computer simulation.

Finally, I discuss further issues that steam from my account, and which are paramount for a fully-fleshed account of explanation for computer simulations. Most prominently is the fact that computer simulations are complex and elaborated systems, and

Entrenching the epistemological side of computer simulations: explanation and unification

as such they might undermine the unificationist approach. I contend this point by offering an alternative approach that takes such complex simulations back to my version of the unificationist account. Another objection stems from focusing my efforts exclusively on equation-based simulation, while the universe of computer simulation is much wider, including cellular automaton and agent based simulations. My response is that each class of computer simulation provides a unique methodology, one that might not be compatible with the unificationist framework. A final issue stems from some conditions imposed on the results of a simulation. Here, I have restricted my view to those results that have an empirical counterpart. It could be the case, however, that the simulation allows results that do not represent empirical phenomena -examples will be discussed. In this context, two questions arise: do we exclude these kinds of results from the study of explanation? or do we find a way to incorporate these cases into our study as well? In this presentation I show how the first question is answered, while I suggest an approach to the second.

References

Batterman, R. W. (2002). *The Devil is in the Details. Asymptoptic Reasoning in Explanation, Reduction, and Emergence*. Oxford University Press.

Bechtel, W. (2005). Explanation: A mechanist alternative. *Studies in History and Philosophy of Biological and Biomedical Sciebces*, 36(2):421–441.

Beisbart, C. (2012). How can computer simulations produce new knowledge? *European Journal for Philosophy of Science*, 2 (3):395–434.

Cohen, K. J. (1961). Two Approaches to Computer Simulation. *Journal of the Academy of Management*, 4(1):43–49.

Craver, C. F. (2006). When mechanistic models explain. 153:355–376.

Durán, J. M. (2013). *Explaining simulated phenomena: A defense of the epistemic power of computer simulations*. PhD thesis, Universität Stuttgart.

El Skaf, R. and Imbert, C. (2012). Unfolding in the empirical sciences: experiments, thought experiments and computer simulations. 190(16):3451–3474.

Freedman, D. and Humphreys, P. W. (1999). Are there algorithms that discover causal structure? 121(1):29–54.

Frigg, R. and Reiss, J. (2009). The philosophy of simulation: Hot new issues or same old stew? 169(3):593–613.

Hartmann, S. (1996). The World as a process: simulations in the natural and social sciences. In Hegselmann, R., Mueller, U., and Troitzsch, K. G., editors, *Modelling and Simulation in the Social Sciences from the Philosophy of Science Point of View*, pages 77–100. Springer.

Humphreys, P. W. (2004). *Extending ourselves: Computational science, empiricism, and scientific method*. Oxford University Press.

Kitcher, P. and Salmon, W. C., editors (1989). *Scientific explanation*. Minnesota Studies in the Philosophy of Science. University of Minnesota Press.

Krohs, U. (2008). How digital computer simulations explain real-world processes. *International Studies in the Philosophy of Science*, 22(3):277–292.

Morgan, M. S. (2005). Experiments versus models: New phenomena, inference and surprise. *Journal of Economic Methodology*, 12(2):317–329.

Morrison, M. (2015). *Reconstructing reality. Models, mathematics, and simulations*. Oxford University Press.

Naylor, T. H., Burdick, D. S., and Sasser, W. E. (1967). Computer simulation experiments with economic systems: the problem of experimental design. *Journal of the American Statistical Association*, 62(320):1315–1337.

Salmon, W. C. (1984). *Scientific explanation and the causal structure of the world*. Princeton University Press.

Weirich, P. (2011). The explanatory power of models and simulations: a philosophical exploration. *Simulation & Gaming*, 42(2):155–176.

Winsberg, E. (2010). *Science in the age of computer simulation*. University of Chicago Press.

Woodward, J. (1989). Data and Phenomena. 79:393–472.

Woodward, J. (2003). *Making things happen*. Oxford University Press.

III Conferencia de Graduados de la SLMFCE

Epistemological disjunctivism as a solution for underdetermination-based skepticism

Eduardo Martínez Zoroa
Universidad de Barcelona

Abstract: I explore a strategy to solve the skeptical argument based on the principle of underdetermination. This strategy consists in the rejection of the premise that the epistemic support which a regular subject has is not enough to favour her ordinary beliefs over radical skeptical hypotheses. I introduce McDowell's and Duncan Pritchard's epistemological disjunctivism as a possible way to reject this premise. Nevertheless, my conclusion is that this anti-skeptical strategy can only work if S's epistemic support is not thought as reflectively accessible, and therefore that epistemological disjunctivism doesn't work.

Key words: skepticism, underdetermination, accessibility, perception, epistemic support, justification, disjunctivism, discrimination.

Radical skeptical hypotheses depict scenarios in which an agent has experiences subjectively indistinguishable from the ones she ordinarily has, but where the agent is massively deceived. These scenarios are used to construe skeptical arguments (which aim to prove that we lack justification or knowledge for most of our beliefs) or skeptical paradoxes (which don't draw any conclusion but just show a putative incompatibility between intuitive claims which concern our epistemic situation).

As an instance of a radical skeptical hypothesis, we are going to use the brain in a vat hypothesis (BIV). According to this hypothesis, one is a brain in a vat instead of a person, but a scientist with a supercomputer is producing experiences in this brain which are identical to those a person would have. Therefore, if S were a BIV S would have experiences subjectively indistinguishable from the ones S would have if S were a normal person. We can use this BIV hypothesis in conjunction with the following epistemic principle in order to build a skeptical argument/paradox:

Underdetermination principle (UP). For all S, p, q, if S's epistemic support for believing p does not favour p over some hypothesis q which S knows to be incompatible with p, then S's epistemic support does not justify S in believing p.

This would be the UP-based skeptical argument:

(1) S's epistemic support for believing p does not favour p over BIV.

(2) If S's epistemic support does not favour believing p over BIV, then S's epistemic support does not justify S in believing p.

(3) S's epistemic support does not justify S in believing p (from 1, 2).

(4) S is not justified in believing p (from 3, and the assumption that justification is based on epistemic support).

(5) S doesn't know p (from 4, and the assumption that justification is a necessary component of knowledge).

And this the UP-based skeptical paradox:

(1) S's epistemic support for believing p does not favour p over BIV.

(2) If S's epistemic support does not favour believing p over BIV, then S's epistemic support does not justify S in believing p.

(3) S is justified in believing p.

An interesting strategy, if one wants to defend that S is justified in believing what she ordinarily does, is trying to reject the first premise of the argument. We want to say that S perceptual experience gives her epistemic support which favours p over BIV. One proposal which aims to sustain this is Pryor's dogmatism (2000, 2004), which says that S's perceptual experience of p makes S's belief that p justified. Perceptual experiences, in virtue of their propositional content (the way they represent the world to be) make S justified in holding certain beliefs. However, non-perceptual, deceptive experiences, can have the very same propositional content that perceptual experiences, as long as they look the same way from S's point of view. Consequently, this epistemic support is defeasible, it doesn't entail that p. S could have an experience as of p while p is not the case, perhaps because he is suffering an illusion or hallucination.

The problem with this account is that it really does nothing to explain how S's defeasible perceptual evidence can favour p over BIV. The point of the skeptic is that the seeming as if p would be exactly the same, and would be equally well explained, if BIV were true instead of p. The content of the seeming as if p doesn't really favour p over BIV because in both cases the content would be exactly the same, or so would say the skeptic. The problem for this kind of dogmatism, then, is that it doesn't provide any story to explain how the first premise of the skeptical argument is false. It takes for granted that the seeming as if p favours p over BIV, but doesn't explain how.

The dogmatist assumes that S's perceptual epistemic support for p is something like S's experience as if p were the case. If the epistemic support is understood as a mere seeming, then it can be replicated in situations where p is false, like the BIV scenario. But one can argue that S's perceptual epistemic support for p does not consist in the experience as if p, but rather in her seeing that p is the case. If the experience is understood this way, then it would not be possible to replicate the epistemic support it provides in situations in which p is false.

The latter one is the strategy followed by McDowell's (1998) and Duncan Pritchard's (2012, 2015) epistemological disjunctivism. This position sustains that, when one is in an optimal case of perception, one has epistemic support which is both factive (having perceptual epistemic support for p entails the truth of p) and reflectively accessible (one can know that this

epistemic support obtains by reflection alone). This epistemic support is not present in bad cases of perception or non-perceptual cases, even though this cases can be subjectively indistinguishable. If we sustain the claim that the content of S's perceptual experience entails that p, then we would have an explanation of why S's perceptual evidence favours p over BIV. Obviously, if the perceptual evidence entails p, and p and BIV are incompatible beliefs, then the perceptual evidence favours p over BIV.

However, according to a broadly accepted assumption, if S's epistemic support is reflectively accessible to S then it is non-entailing, because S's reflectively accessible epistemic support is thought to be compatible with the target belief being false. The argument can be casted as follows:

(1) In the bad case, S's reflectively accessible epistemic support is less than factive.

(2) One cannot reflectively distinguish between a good case and a corresponding bad case.

(3) In the good case, the reflectively accessible epistemic support is less than factive.

McDowell's strategy against this argument is to point out the difference between accessibilism (the thesis that S's epistemic support must be reflectively accessible) and what he calls the Highest Common Factor conception (HCF).

HCF: The only facts that S can know by reflection alone in an optimal case of perception are facts that S's physical duplicate in a corresponding bad case can also know by reflection alone (by "corresponding bad case" is meant that everything looks to S as if she were in a good case of perceptual, but S is actually suffering a deceptive experience).

McDowell argues that a commitment to accessibilism doesn't entail a commitment to HCF, because the fact that both experiences are indistinguishable from S's point of view doesn't entail that both experiences provide the same epistemic support. According to McDowell, it's not a commitment to accessibilism, but an independent commitment to HCF, what prevents philosophers to say sustain that S's reflectively accessible epistemic support can be factive. The most that will follow from the fact that the two experiences are indistinguishable is, according to McDowell, that the agents in the good and in the bad case are equally epistemically blameless in having the target belief (this would be the only internalist epistemic standing which both agents have in common). But this is not the same as saying that both agents have the same epistemic support reflectively available or that both are equally justified in believing the target proposition.

Nevertheless, McDowell doesn't tell any story to explain how can the epistemic support be different in the two cases if the experiences are indistinguishable. An attempt to answer this worry can be found in Pritchard's defence of epistemological disjunctivism (2012). Pritchard's strategy is to claim that, contrary to what it appears at first sight, it is not true that, when the experiences look the same to S, S cannot distinguish between the good and the bad case on the basis of her reflectively accessible epistemic support. So he aims to reject (2) "One cannot reflectively distinguish between a good case and a corresponding bad case", and then stop the inference to (3) "In the good case, the reflectively accessible epistemic support is less than factive.". According to Pritchard, the assumption that S cannot reflectively distinguish between the two cases trades on an ambiguity of the idea that the two scenarios are reflectively indistinguishable. He defends that there are two senses of "reflectively distinguishing" which should be separated.

1) One sense is that of *perceptually discriminate* one scenario from the other. For instance, imagine that S were a zoologist trained so that she could be perceptually aware of certain differences between genuine zebras and disguised mules. In this case, S's discrimination is based on the way things look to S. S distinguishes that there is a zebra (or a disguised mule) "just by looking" (Pritchard, 2012, p. 77), because her visual experience is enough for her to discriminate between the two scenarios.

2) Another sense in which S can reflectively distinguish between the two scenarios is when S's *background epistemic support favours* S's belief that she is in a good rather than a bad scenario. Imagine that S is not a zoologist and she has no special ability to perceptually discriminate between zebras and disguised mules. The two experiences are indistinguishable from S's point of view. Still, it could be argued that S's background epistemic support favours the claim that she is in the good case, in particular what she knows about the low-likelihood of a deception of this kind taking place (she could reason that it is a pointless deception, and that it involves a huge effort, that the zoo is a reliable one, etc). So, even though she is unable to introspectively discriminate between the two scenarios on the basis of her visual experience, S can offer favouring epistemic support for her belief that she is in the good scenario over alternative hypotheses, via a priori deduction based on her background rational beliefs. Then, according to Pritchard, we can say that S can reflectively distinguish whether or not she is in a good case on the basis of her background epistemic support and, consequently, reject premise (2) of the distinguishability argument against epistemological disjunctivism.

This distinction between favouring and discriminating epistemic support is not distinctive of epistemological disjunctivism, and could be accepted independently of it. But, according to Pritchard, it helps epistemological disjunctivism because it allows one to claim that S can favour p over incompatible hypothesis on the basis of reflectively accessible evidence, even though the way things looks to S is not enough to favour p over q. What is reflectively accessible to S when it comes to favour one perceptual belief or the other is not only what S can discriminate on the basis of how things look to her, but also her background of reflectively accessible justified beliefs, which can be used as further evidence to favour one scenario over the other.

Epistemological disjunctivism as a solution for underdetermination-based skepticism

But I think that Pritchard's argument doesn't work, as long as he also wants to sustain that perceptual evidence is factive and reflectively accessible. If we accept the claim of epistemological disjunctivism that S's perceptual evidence is factive, then this support cannot be exhausted neither by her discriminating nor by her favouring epistemic support for p. Neither the kind of epistemic support provided by the favouring background of S's beliefs nor the kind of epistemic support provided by what S is able to discriminate from how the experience looks to her is enough to provide S with factive epistemic support. The epistemic support provided by what is accessible in the visual experience is compatible with scenarios in which the claim is false, as the distinguishability argument shows. The epistemic support provided by S's background beliefs can make a belief more likely, but it is not enough to entail it. So, if it is admitted that what is reflectively accessible to S is what she can distinguish on the basis of how things look to her and what she can favour on the basis of her justified background beliefs (and this seems to be the epistemic support which Pritchard acknowledges to be reflectively accessible), then neither of these can play the role of the factive epistemic support which epistemological disjunctivism needs.

This is why epistemological disjunctivism must say that there is something more which is relevant to S's epistemic position, besides what she can distinguish on the basis of her experiences and beliefs. This "extra", factive, epistemic support is thought to be provided by the relation between S and the relevant fact of the world. That is, it's not what S can distinguish from that experience, but the fact that this experience involves a relation with facts of the world, what provides S with factive epistemic support.

> The particular kind of rational support that the epistemological disjunctivist claims that our beliefs enjoy in paradigm cases of perceptual knowledge is that provided by seeing that the target proposition obtains … [In the good case] the agent's reflectively accessible rational support will be the factive ground that there is a tree before her. (Pritchard, 2012, pp. 14-16)

But then one could argue, against Pritchard's argument for epistemological disjunctivism, that S cannot distinguish whether she is in such a relation with the relevant fact of the world, because the means which she has for it (what she can distinguish on the basis of the experience and what she can favour on the basis of her background beliefs) are not enough to determine whether she is actually in a relation with the relevant fact or she is in a bad case instead. In other words, if, in a case of perception, the factive epistemic support is provided by the relation between S and some external fact, then the epistemic support is not reflectively accessible to S, because the occurrence of this relation is beyond S's accessible reach (the causal relation between the experience of p and the fact that p is beyond what is accessible from the experience).

It seems then that we must choose between the claim that S's perceptual epistemic support is factive and the claim that S's perceptual epistemic support is reflectively accessible. The interesting feature of epistemological disjunctivism is that, by sustaining both claims at the same time, it allows a defence of the claim that S's epistemic evidence is reflectively accessible but capable of favouring p over skeptical alternatives. But we have concluded that the combination of these two claims doesn't seem to be possible.

References

McDowell, J. (1998). "Criteria, Defeasibility and Knowledge", In *Meaning, Knowledge, and Reality*, Harvard University Press.

Pritchard, D. (2012) "Epistemological Disjunctivism", Oxford University Press, Oxford, 2012.

——— (2015) "Epistemic Angst: Radical Skepticism and the Groundless of our Believing", Princetown University Press.

Pryor, J. (2000) "The Skeptic and the Dogmatist", Noûs, 34, 517-49.

——— (2004) "What's Wrong with Moore's Argument?" in *Philosophical Issues,* 14, 349-378.

III Conferencia de Graduados de la SLMFCE

Relevance Theory, grammar and processing effort: how grammar diversity affects the explicitness of utterances

Joan Gimeno Simó
University of València

Abstract:
This proposal aims to lay out a problem concerning the approach of Relevance Theory to explicit communication. The main idea that we will argue for is that this theory is unable to deal with the fact that the grammar of some languages may oblige its speakers to make more explicit some items that would remain unmarked in other languages, thus making their utterances costlier to process without increasing their overall relevance. If we are right about this, grammar may play a bigger role than optimal relevance in determining the degree of explicitness of utterances.

Keywords: Relevance Theory, explicitness, grammar

Resumen:
El presente artículo pretende presentar un problema relativo a la aproximación a la comunicación explícita por parte de la Teoría de la Relevancia. La idea principal que sostendremos es que dicha teoría resulta incapaz de tratar con el hecho de que la gramática de algunas lenguas puede obligar a sus hablantes a hacer más explícitos algunos elementos que permanecerían sin marcar en otras lenguas, aumentando el esfuerzo de procesamiento requerido por sus enunciaciones sin que esto aumente su relevancia. Si estamos en lo cierto, la gramática podría jugar un papel más importante que la relevancia óptima a la hora de determinar el grado de explicitud de las enunciaciones.

Palabras clave: Teoría de la Relevancia, explicitud, gramática

Our proposal attempts to lay out what we think to be a problem for the approach of Relevance Theory to explicit communication. Relevance Theory takes relevance and processing effort as the main contributors to the determining of the degree of explicitness of an utterance; we will argue that, while these factors are important, sometimes grammar may play a bigger role in this process.

We will start out by making a brief introduction to those features of Relevance Theory that concern the most to our aims. This approach differs from Gricean approaches regarding explicit communication: Gricean frameworks considered that grasping the explicit meaning of an utterance requires just disambiguation and reference assignment, while Relevance Theory takes what is explicitly communicated as going much beyond what is literally said. Thus, Gricean theories consider pragmatic inference as a means to recover what is implicated, while Relevance Theory sees it as not just contributing to the recovery of the implicatures, but also to the development of what relevance theorists call the explicature of the utterance: once the utterance is decoded, the output of this decoding must be developed until the hearer gets a full propositional form. This means that, while Gricean theories consider what is explicitly said as equating what is literally said, relevance theorists prefer to talk about *degrees of explicitness* of utterances: the greater the contribution of the decoding and the less that of the pragmatic inference, the more explicit the utterance is.

This thesis from Relevance Theory is sustained on the basis that utterances follow the Principle of Relevance: any act of ostensive communication – such as an utterance – communicates the presumption of its own *optimal relevance*. An stimulus is optimally relevant if it leads the hearer to infer the maximum number of relevant assumptions with the less *processing effort*. Thus, given two different utterances that allow the hearer to make the hearer infer the same set of relevant assumptions, a speaker willing to communicate that very set of assumptions will choose the one that takes the less effort to process. The hearer is supposed to decode this utterance and to develop the output of this decoding process until she gets an interpretation fitting the Principle of Relevance. The result of this development is what relevance theorists call an *explicature*. Let us illustrate this by means of the following example:

(1a) "That book is difficult"
(1b) "That book is difficult to read"

In a context in which it is clear that we are talking about reading books, a speaker willing to communicate a certain set of assumptions will prefer to utter (1a) over (1b), since it takes less effort to process; the hearer is supposed to decode it and to develop the output of this decoding process until she gets a proper interpretation – in this case, that the book is difficult *to read.* Given two utterances that communicate the same set of assumptions, a speaker aiming at optimal relevance will generally choose the syntactically simpler one, since it would be less costly (Sperber and Wilson, 2012: 377; 1993: 9).

In what follows, we will proceed to show how this idea from Relevance Theory fails, and we will do it by showing how the information encoded by speakers in order to communicate a set of assumptions varies depending on the language that speakers are using. We will argue that the extra processing effort required by the occurrence of some items cannot be justified by their relevance, but only because of grammar conventions.

Relevance Theory, grammar and processing effort: how grammar diversity affects the explicitness of utterances

Our methodology will be as follows: we will imagine several situations in which an utterer wants to communicate a certain set of assumptions, and we will see how the information encoded varies depending on the language. We will see how there are some cases that Relevance Theory can explain away by appealing to lexical differences, while there are others in which this explanation just does not hold, and we must appeal to grammatical rules.

Slobin (2011) has shown some cases in which a speaker intending to communicate a set of assumptions by means of an utterance encodes different information depending on which language he is using to communicate it. Namely, he argues that some languages tend to encode pieces of information that are left to be inferred by speakers of other languages. Thus, for example, he has shown that Spanish speakers tend to encode the final state of an action, leaving the direction of the action to be inferred by the hearer, whereas English speakers tend to encode the direction of the action, leaving its final state to be inferred.

According to relevance theory, a speaker willing to communicate a set of assumptions by means of an utterance will only encode the minimum information required for the hearer to grasp the intended meaning, since encoding more information would make the utterance costlier to process, and thus *non-optimally relevant*. If this is so, we may ask ourselves why an utterer should encode more information in a language *a* than in a language *b*: if both utterances convey the same explicatures, and the utterance in language *b* is sufficient for the hearer to infer the intended set of assumptions, the extra information provided when uttering that very sentence in language *a* seems unnecessary. Let us consider the following example: a kid is eating an ice-cream and gets some stains on his clothes. His father sees him and tells him to wash the stains off. This is how the utterance would have gone in English and in Spanish:

(2a) "Wash that off!"
(2b) "¡Lávate eso! ("Wash *to yourself* that")

We can see from this example that the information encoded by each utterance differs. (2a) is encoding, by means of the particle "*off*", the fact that the hearer has to wash something *off* a surface, whereas (2b), by means of the reflexive "*-te*", is encoding the fact that the hearer has to perform the action of washing on himself. Neither of these pieces of information were required: a Spanish hearer would not require the speaker to encode "*de encima*" ("from the surface") to grasp the intended meaning – it was already in the explicature of (2b) –, and neither would an English hearer need that the speaker encoded "to yourself" to grasp the fact that he must perform that action on himself (it was already in the explicature of (2a)). What is, then, the reason for encoding such information?

The answer that relevance theorists could provide to the problem that we have laid out may be the appeal to lexical differences: language *a* may have a word encoding a concept more precise than the word used by the speaker when uttering it in language *b*. Thus, for example, a German speaker may use the verb "*kennenlernen*" (paraphrasable as "to get to know") whereas an English speaker would be using the less precise "to know". Decoding two verbs is costlier than decoding a single one, hence in those cases in which no ambiguity rises an English speaker aiming at optimal relevance would prefer to translate "*kennenlernen*" as "to know" instead of as "to get to know". On the other hand, German has two distinct verbs, "*kennenlernen*" ("to get to know") and "*kennen*" ("to know"). Since both of them require the same processing effort, the German speaker will tend to use the one that is closer to the intended meaning. This allows us to understand why the German sentence may be more precise than its English counterpart: each speaker is just using the best resources that her language provides her with, keeping the processing effort as low as possible. The Principle of Relevance remains therefore untouched.

This response requires the adoption of a contextualist framework in which not all concepts are lexicalized, meanings of utterances are built *ad hoc* for each situation, and words are used as a simple clue for grasping the actual meaning of the speaker, since they encode only loose concepts. This response allows us to explain away differences between, for example, the degrees of explicitness of (2a) and (2b). (2a) is employing "to wash off" as a verb, whereas (2b) is using "*lavarse*" ("to wash oneself"). Neither of these verbs has a perfect equivalent in the other language, but both of them require the same processing effort as the less precise "to wash" and "*lavar*" ("to wash"); the reason for the choice of the more precise one is that speakers tend to use those words that are closer to the intended meaning. The fact that they are encoding unnecessary information does not affect the optimal relevance of the utterance, since processing either "Wash that off!" or "Wash that!" requires the same effort, and the same goes for "*¡Lava eso!*" ("Wash that!") and "*¡Lávate eso!*". The difference, then, is just lexical: each speaker is just using the best resources that her language allows to, and it happens that a language has lexicalized a concept that the other has not.

We think that this response is right; however, in our opinion, it still fails to provide a proper answer to other similar problem that cannot be explained away in terms of lexical differences. Sometimes it is the grammar of some languages, and not their lexical features, that obliges their speakers to provide an extra information that is not required in other languages to grasp the set of utterances that the speaker is intending to communicate. It is the case

Relevance Theory, grammar and processing effort: how grammar diversity affects the explicitness of utterances

of Chinese sentences containing the particle 了 ("*le*"; a particle employed to indicate a change of state) or sentences containing some pronouns found in several Romance languages (French "*y*" and "*en*", Italian "*ci*" and "*ne*", Catalan "*hi*" and "*en*"; used as anaphoric pronouns to substitute an adverbial phrase or a noun modifier). Utterances employing these items are more explicit than their counterparts in other languages where these particles do not exist or are not compulsory, since the occurrence of these particles implies encoding some truth-conditional (non-illocutionary) information. This extra encoding seems unnecessary, since this information was already in the explicature of the utterance and the speakers of other languages did not need to encode it for the hearer to grasp what was intended to communicate. Compare the following conversarion in English ((3a) and (3b)) and its translation into Catalan ((3c) and (3d)):

(3a) "Why did you leave the party?"
(3b) "I didn't feel comfortable
(3c) "Per què vas deixar la festa?" ("Why did you leave the party?")
(3d) "No m'*hi* trobava còmode" ("I didn't feel comfortable *at the party*")

The encoding of the particle *hi* in (3d) is making this sentence more explicit than its English counterpart (3b). This pronoun is encoding the same propositional information as the italicized expression in the translation of (3d), that is, it is substituting the adverbial phrase "*at the party*". The occurrence of this pronoun is compulsory, and therefore obliges the Catalan speaker to make the sentence syntactically more complex than it could have been if the speaker had chosen to encode just the information encoded in its English counterpart (3b), where the adverbial phrase "*at the party*" was absent. This extra information, therefore, has not been added because it was relevant in the given context: its occurrence, and the additional effort to process it, can only be justified by grammar.

This time the problem cannot be dealt with in terms of lexical differences, since it would require accepting that the combination of "*hi*" plus a verb is just a lexicalization of a new, more precise concept. This would eventually lead us to consider all the combinations of verbs plus pronouns as encoding concepts that are as primitive as those encoded by verbs alone, which seems rather implausible: we should consider, for example, "to see her" as being as primitive as "to see". Even if this were so, this would not explain the extra effort required for processing it, since, unlike in the case of "*kennenlernen*" and "to know", which are each one encoding a single verb, "to see" is encoding a verb, whereas "to see her" is encoding both a verb and an object, which would make it costlier to process.

Let us see a second example, this time comparing English, Spanish and Chinese. Imagine a speaker inviting the hearer to enter a room. If the speaker is inside the room, the utterance would go like this:

(4a) "Come in!"
(4b) "¡Entra!" ("Enter!")
(4c) "进来！" （"Jìn lái!) ("Enter come!")

The English utterer of (4a) is employing a single verb that indicates the direction of the action (*towards the speaker*) and the final state of it (the hearer must en up *inside*). The Spanish utterer of (4b) is employing a single verb that encodes that the hearer has to perform a movement by which he will end up *inside*. The Chinese utterer of (4c) is using two different verbs, the first one to indicate that the movement must end up *inside* and the second one to indicate the direction of the action *(towards the speaker*). This means that the Chinese and the English speaker are encoding the same information, whereas the Spanish utterance lacks the direction in which the action must be performed. The difference between the degrees of explicitness of (4a) and (4b) can be explained away by appealing to lexical differences, since the syntax of each of them is as simple as it could have been. But one may wonder why the Chinese speaker is using two different verbs: the information encoded in the Spanish utterance would have sufficed for the hearer to grasp the intended meaning, and there was no need to incorporate a second verb indicating the direction of the action. Therefore the relevance of this information cannot justify its occurrence nor the extra effort required to process it: it can only be due to grammar.

According to Relevance Theory, the main reason to make an utterance more explicit than another one conveying the same explicature – that is, the reason to encode more information – is to indicate the direction in which the hearer has to search for relevance, that is, to provide the hearer with a clue of what she has to do with the utterance; this is the only thing that can justify the extra effort required for processing it. However, having the above problems in mind, we can conclude that an utterance's being more explicit than another one may also depend on grammar conventions. Indeed, in the cases shown above we can see how speakers may sacrifice the optimal relevance of an utterance in order to follow a grammatical rule: by chosing to follow it, the utterance is made costlier than it could have been.

Relevance Theory, to a great extent, dismissed semantics in favor of pragmatics, since it attributed the latter a role not just in the study of what is implicated, but also in what is explicitly said – the explicatures. However, if we are right in our claims, the role of pragmatic inference in the recovery of explicatures may not be as great as relevance theorists take it to be, since the degree of explicitness of utterances may be more determined by grammar than by the optimality of the relevance of utterances.

References

BLASS, R. (1989) "Pragmatic effects of co-ordination: the case of 'and' in Sissala". *UCL Working Papers in Linguistics*. University

College London.

CARSTON, R. (2011) "Relevance theory", in G. Russell & D. Graff Fara (eds.), *Routledge Companion to the Philosophy of Language*. Routledge.

CARSTON, R. (2002) *Thoughts and Utterances: the Pragmatics of Explicit Communication*, Malden (MA): Blackwell.

GRICE, H. P. (1989) *Studies in the Way of Words*. Harvard University Press, Cambridge MA.

RECANATI, F. (2004) *Literal Meaning*, Cambridge University Press, Cambridge.

SLOBIN, D. I. (2011) "Thinking for speaking", *Proceedings of the Annual Meeting of the Berkeley Linguistics Society* (Vol. 13), pp. 71-96.

SPERBER, D. AND WILSON, D. (1993) "Linguistic Form and Relevance", *Lingua* 90, pp. 1-25.

SPERBER, D. AND WILSON, D. (2012) "Pragmatics", *Meaning and Relevance*, Cambridge: Cambridge University Press, pp. 1-27.

SPERBER, D. AND WILSON, D. (1986) *Relevance. Communication and Cognition.* Oxford: Blackwell.

SPERBER, D. AND WILSON, D. (1997) "The Mapping between the Mental and the Public Lexicon", *UCL Working Papers on Linguistics* 9.

VALOR ABAD, J. (2015) "Les Paradoxes i Filosofia: tres Visions Contemporànies", *Quaderns de Filosofia (2)*, vol. 2, pp. 67-98.

WRIGHT, C. (1976), "Language Mastery and the Sorites Paradox", Evans, G. and McDowell, J. (1999) *Truth and Meaning. Essays on Semantics*, Oxford: Oxford University Press, pp. 223-47.

III Conferencia de Graduados de la SLMFCE

The causal structure of evolutionary theory: the scope and limits of the force interpretation

Victor J. Luque
Evolutionary Genetics Group
Cavanilles Institute of Biodiversity and Evolutionary Biology

ABSTRACT: This article analyzes the view of evolutionary theory as a theory of forces. The analogy with Newtonian mechanics has been challenged due to the alleged mismatch between drift and the other evolutionary forces. Several authors have postulated that the special character of drift is because it is the default behaviour or Zero-Cause Law of evolutionary systems. I defend that drift's causal and explanatory power prevents it from being considered as a Zero-Cause Law. Instead, I propose that the default behaviour of evolutionary systems is what I call the Principle of Stasis. This approach has several advantages in order to defend the causal status of evolutionary theory and explains the use of the force interpretation.

KEYWORDS: Evolutionary forces, causation, Zero-cause law, genetic drift, Newtonian analogy.

Textbooks and most of the evolutionary literature talk about e*volutionary forces* acting on a population. The analogy was proposed by Elliott Sober (1984) who argues that evolutionary theory is a theory of forces because, in the same way that different forces of Newtonian mechanics cause changes in the movement of bodies, evolutionary forces cause changes in gene and/or genotype frequencies. As a result, selection, drift, mutation and migration would be the main forces or causes of evolution. Nevertheless, the appropriateness of the causal view, and particularly the Newtonian analogy, has been challenged in the last decade. Several authors (Walsh *et al* 2002, Matthen and Ariew 2002) argued for a new view, the *statistical view*, where the evolutionary process and its parts (selection, drift, etc.) are mere statistical outcomes, inseparable from each other.

I argue for a third way to defend the causal view. The aim of the force interpretation was to expose the causal structure of the theory. This is what Maudlin (2004) calls "quasi-Newtonian" theories. These are characterized by shaping them into a similar form to Newtonian mechanics whose main axis is the adoption of a default behaviour which tells us how the system would behave if external factors were not acting on it. I call Zero-Cause Law (henceforth ZCL) this default behaviour. The main purpose of building quasi-Newtonian theories is to identify the causes that affect a particular system. That is why the ZCL is necessary.

Several authors still think that drift is an evolutionary force inasmuch as it is an essential causal factor in evolutionary phenomena. Stephens (2004) postulates that drift has a direction: it leads populations to homozygosity. Filler (2009) is in favour of maintaining the *force-talk* but is aware that abuse of it could turn the concept of force into nonsense. Pence (2016), on the other hand, finds in Brownian motion, in physics, a similar phenomenon –a stochastic force– so we can maintain drift as a force. Other authors give to genetic drift a special role which could be explained considering it not as a force but as a "default state". McShea and Brandon (2010) defend that drift is the ZCL instead of the Hardy-Weinberg law (H-W law). McShea and Brandon point out the difficulties of considering drift as a force because of its lack of direction. Instead, these authors defend that drift, far from being a special force which is introduced in the population, is the default state of the population –because there is no infinite populations– and, hence, a ZCL in the same way that the inertia is bodies' default state in Newtonian mechanics. In a similar line of argumentation, Sarkar (2011) locates drift as ZCL but with different connotations from those defended by McShea and Brandon. Sarkar build a haploid model and points out that drift is not mentioned in the model. However, it is included in the model through the population size when it is finite. Population size actually is a *constitutive assumption* of the system. Constitutive assumptions are those privileged conditions which cannot be changed without changing the identity of the system. Facultative assumptions, on the other hand, are those which may vary without changing the identity of the system.

A critical analysis of the role played by drift within the structure of evolutionary theory will show the scope and limits of the force interpretation (for more details see Luque (2016) and Luque (in press)). The force interpretation was proposed to help identify evolutionary causes. Nevertheless a theory can be a causal theory without resorting to forces. I argue for a difference-making account of causation (Menzies 2004). According to this approach, then, a cause is conceptualized as a *difference-maker*, disturbing the normal behaviour of the system. In other words, a cause is "what makes the difference in relation to some assumed background or causal field" (Mackie 1980, p. xi). The system is defined by a number of background conditions, and among these conditions the ZCL tells us how the system behaves before the intervention of external factors, what the *normal course* of the system is like. Some authors (McShea and Brandon 2010) call a *default state* the normal course of the system. However, I think that *default behaviour* is preferable

The causal structure of evolutionary theory: the scope and limits of the force interpretation

because a default state of a system is shaped not only by the ZCL, but also by other default settings or background conditions. Thus, difference-making factors "are seen as intrusions into the system that account for the deviation from the normal course of events" (Menzies 2004, p. 170).

This kind of theorizing is found in Population Genetics textbooks by, firstly, establishing the background conditions of the system and, secondly, by introducing factors against this background. Evolutionary theory usually takes for granted the Hardy-Weinberg law (henceforth H-W law) (Sober 1984, Gillespie 2004) as its ZCL counterpart. According to the H-W law a diploid and ideal infinite population, where there is random mating (panmictic population) and whose individuals are viable and fertile, will remain or return to equilibrium (i.e. gene and genotype frequencies will remain stable) if no external factor acts on it. The best historical example following this way of theorizing is Newtonian mechanics (Menzies 2004, Maudlin 2004). Thus, the first law of Newtonian mechanics functions to establish that every body continues in its state of rest, or of uniform motion in a right line, unless it is compelled to change that state by forces impressed upon it. Thus, both the law of inertia as well as the H-W law, tell us how the system would behave if nothing disturbed it, and so assuring a neutral substrate where we can introduce external factors.

Within the background conditions, ZCLs play a crucial role because they tell us how the system behaves before the intervention of external factors and what the normal course of the system is like. Since the ZCLs are part of the background conditions, they are the basis to explain the unexpected. They are what is expected (the default behaviour), but because of that, they are required to identify what must be explained. That is the role that Newton gives to the law of inertia. It tells us that when no force is operating, the body will continue at a constant velocity. In addition, Newton's First Law is the link between the background conditions and difference-makers in any Newtonian system, since it puts us in a position to appreciate the effects of different forces and is the only background condition which plays that role. The law of inertia presupposes the following features: absolute space, absolute time and the existence of one body (Maudlin 2012). The first two give us a topology, a structure, a metric and a vacuum where the body and its state of motion can be located, measured and continued in a straight line (if it is not at rest). At the same time, the law of inertia requires the inclusion of external forces (difference-makers) in order to explain why a body is not in uniform motion (in a straight line or at rest).

I defend that theoretical and empirical reasons and drift's causal and explanatory power prevents us from considering it as a ZCL, because it does not correspond to the features of ZCLs. Universality plays in favour of considering drift as a ZCL. But this feature is not enough. Think about the force of gravity. Newton formulated it as a universal law and, in fact, it operates anywhere in the universe where bodies are interacting. We could say that it is a constituent part of the universe, its default behaviour and because of that, it should be considered a ZCL.

Following Sarkar's distinction between constitutive and facultative assumptions, we can see why drift is not the default behaviour. The distinction implies identifying the features that cannot be removed without changing the essence or the identity of the system. I propose that the default settings of evolutionary theory (at least in Population Genetics) can be summarized as follows: a population, variation, an environment, ancestor/descendant relations, and what I call The Principle of Stasis. Evolution requires a population of individuals because individuals, themselves, do not evolve; only populations evolve. Evolution also requires non-identical individuals, the existence of variation in a population. The environment is where the population is located, the action space where the population develops its activities. Ancestor/Descendant relations give to the population a time line and it is not committed to any particular form of heredity. The Principle of Stasis tells us how the system will behave if there are no facultative assumptions. It is the ZCL that connects constitutive and facultative assumptions. We can see that the Principle of Stasis depends on the Ancestor/Descendant relations, in the same way that the Principle of Inertia depends on absolute space (the vacuum), or the existence of a body.

On the other hand, facultative assumptions (i.e. evolutionary causes or difference-makers) can be removed without changing the identity of the system. We could devise a prototypical standard evolutionary theory system without natural selection by postulating no fitness variation. Also we could think of an evolutionary system without mutation by postulating a perfect replication mechanism or without migration by postulating a closed population. In all these cases, the absence of the processes does not affect the nature of the system. And, finally, the mathematical models in Population Genetics allow us to see why drift is not a background condition. In any Population Genetics textbook drift is introduced by postulating a finite population size. In order to make the calculus easier, models start with an infinite population to construct a deterministic process. Thus, we can perfectly model an evolutionary system with an infinite population size and the system will remain within standard evolutionary theory's framework. What it shows is that we can model an evolutionary system without drift. What is a background condition is the population itself and not its size.

Another problem that appears when we consider drift as a ZCL is that it has a central role in the explanation of a large number of evolutionary phenomena**.** Lynch (2007) has suggested that the increase in genome size occurred in the transition from prokaryotes organisms to eukaryotes organisms has been due to genetic drift which fixed in genomes, along lineages, elements with little or no advantage, and even mildly deleterious, such as introns, transposable elements, noncoding DNA, etc. Drift would be the main factor because the eukaryotes effective population size was smaller than the prokaryotes one.

Thus, bacterial species have an effective population size so large that selection quickly fixed any beneficial mutation and its ability to eliminate the deleterious one is huge. On the other hand, in eukaryotic species, the effective population size is much smaller so that natural selection is not as effective in eliminating those elements. Thus, drift is shown as a crucial factor in explaining the increase of organisms' complexity. However, its important explanatory and causal role is what makes it impossible to understand drift as a ZCL. ZCLs are not the cause of anything, but they provide a sort of framework which stipulates what needs to be explained and what is a cause in the system.

I consider the H-W law –in its two possible formulations: allelic and genotypic– as special cases of a more general principle –the *Principle of Stasis*– which could be formulated as follows:

Principle of Stasis: An evolutionary system where there is no selection, drift, mutation, migration, etc., so there is no difference-maker, will not undergo any change (it will remain in stasis).

The structure of evolutionary theory as a quasi-Newtonian theory involves the establishment of a ZCL or default behaviour which indicates how the system would behave if there are no difference-makers. Several authors have claimed the special character of drift within evolutionary theory, postulating it as the ZCL of evolutionary systems. It has been shown that such approach is not well-grounded since drift is not a good ZCL. Instead, I propose a ZCL which includes the H-W law, the Principle of Stasis, which postulates that an evolutionary system, if it is not influenced by a difference-maker, will remain unchanged.

There are several conceptions of the causalist interpretation of evolutionary theory, some of them committed to the Newtonian analogy. My approach has several advantages in order to defend the causalist view. Firstly, the difference-making account allow us to avoid the problems associated with the extension of the concept of "force" beyond classical mechanics; this account does not require that a cause has an explicit and predictable directionality, only that it makes a difference, so drift can still be considered an evolutionary cause without dealing with all the properties that a Newtonian force must have. Secondly, this difference-making account of causation encompasses different causalists approaches. The vast majority have followed Woodward (2003)'s manipulationist account of causation, while others have followed a counterfactual account of causation or a probabilistic account of causation. Difference-making is a general form of all these accounts of causation (manipulationist, counterfactual and probabilistic).

Finally, my approach explains the use of the force interpretation. The Newtonian analogy is illuminating insofar as it is helpful in revealing the causal structure of evolutionary theory. In other words, the theory is constructed from a ZCL that stipulates a default behaviour and arises by introducing factors which alters that behaviour. That is the reason why the force metaphor was formulated in the first place and why it still continues in evolutionary literature.

REFERENCES

Gillespie, J. (2004). *Population Genetics. A concise guide (second edition)*. Baltimore: The John Hopkins University Press.

Luque, V.J. (2016). The Principle of Stasis. Why drift is not a Zero-Cause Law. *Studies in History and Philosophy of Biological and Biomedical Sciences,* 57, pp. 71-79.

Luque, V.J. (in press) Drift and evolutionary forces: scrutinizing the Newtonian analogy. *Theoria,* 31 (3).

Lynch, M. (2007). *The Origins of Genome Architecture*. Sunderland: Sinnauer.

Mackie, J.L. (1980). *The Cement of the Universe (paperback edition)*. Oxford: Clarendon Press.

Matthen, M. & Ariew, A. (2002). Two ways of thinking about fitness and natural selection. *Journal of Philosophy*, 99, 55–83.

Maudlin, T. (2004). Causation, counterfactuals, and the third factor. In: J.D. Collins, N. Hall, L.A. Paul (Eds) *Causation and counterfactuals* (pp. 419–443). Massachusetts: The MIT Press.

Maudlin, T. (2012). *Philosophy of Physics: Space and Time*. Princeton: Princeton University Press.

McShea, D. & Brandon, R. (2010). *Biology's first law: the tendency for diversity and complexity to increase in evolutionary systems*. Chicago: The University of Chicago Press.

Menzies, P. (2004). Difference-making in context. In: Collins JD, Hall N, Paul LA (eds) *Causation and counterfactuals* (pp. 139–180). Massachusetts: The MIT Press.

Pence, Ch. (2016). Is Genetic Drift a Force? *Synthese* DOI 10.1007/s11229-016-1031-

Sarkar, S. (2011). Drift and the causes of evolution. In: P. Mckay, F. Russo, J. Williamson (Eds) *Causality in the sciences* (pp. 444–469). Oxford: Oxford University Press.

Sober, E. (1984). *The Nature of Selection*. Chicago: University of Chicago Press.

Stephens, Ch. (2004). Selection, drift, and the 'forces' of evolution. *Philosophy of Science*, 71, 550–570.

Walsh, D., Lewens, T. & Ariew, A. (2002). The trials of life—natural selection and random drift. *Philosophy of Science*, 69, 452–473.

Woodward, J. (2003). *Making things happen*. New York: Oxford University Press.

III Conferencia de Graduados de la SLMFCE

Polisemia e infraespecificación semántica

Marina Ortega-Andrés
Universidad del País Vasco
marina.ortega@ehu.eus

El objetivo de estas páginas es discutir las soluciones que algunas teorías proponen al problema de cómo codificamos los significados de palabras polisémicas y homónimas. Se estudian tres teorías: la teoría de la selección de sentidos, la teoría de la relevancia sobre los conceptos y la teoría del léxico generativo. Como alternativa a estas propuestas, se plantea que las palabras no codifican conceptos atómicos sino complejos. La estructura del trabajo es la que sigue: primero, explico brevemente la distinción entre homonimia y polisemia; segundo, discuto las tres teorías sobre la polisemia; tercero, propongo que los distintos significados polisémicos se codifican dentro de un mismo concepto complejo.
Palabras clave: polisemia, homonimia, conceptos, significado.

Polisemia y homonimia

La homonimia y la polisemia se caracterizan porque varios significados caen bajo una misma forma léxica. La diferencia entre los dos fenómenos es que en el caso de la polisemia, los dos sentidos están relacionados, en cambio, los significados homónimos no están relacionados entre sí. Observemos el siguiente ejemplo:

(1)

(1a) Pedro esperó a Carlos en el banco.

(1b) De camino, Carlos se pasó por el banco a sacar dinero.

(1c) Carlos tuvo problemas con el director del banco.

En (1), tenemos una palabra "banco" con varios significados. El significado de "banco" en (1a) es completamente distinto al que tiene en (1b) y (1c). En cambio, los significados o sentidos de "banco" en (1b) y (1c) están relacionados de una forma que no se observa en (1a).

Sean:

$Banco_1$, el significado de "banco" en (1a).

$Banco_2$, el significado de "banco" en (1b)

$Banco_3$, el significado de "banco" en (1c)

$Banco_2$ y $banco_3$ son significados polisémicos y $banco_1$ es un significado homómimo con respecto a $banco_2$ y $banco_3$. $Banco_2$ y $banco_3$ están relacionados de una forma que no lo está $banco_1$.

La diferencia entre la homonimia y la polisemia no se reduce a esta similitud intuitiva entre los usos de la expresión. Hay estudios empíricos (Frisson 2009; Kleposniotou et al. 2008; Pylkkänen 2006) que sugieren que existe una diferencia en términos de codificación y procesamiento entre los significados polisémicos y los significados homónimos de una misma palabra. Estos estudios sostienen que en el procesamiento de significados homónimos hay competitividad: los dos significados de la palabra compiten por ser activados durante la interpretación de la palabra. En cambio, en el caso de la polisemia, no hay competitividad, los significados polisémicos no compiten por ser interpretados, sino que facilitan la interpretación de otro significado. De esto se concluye que los significados homónimos y polisémicos no se codifican de la misma forma. Mientras que los significados homónimos parecen codificarse en paralelo, los significados polisémicos no lo hacen.

Explicaciones de la polisemia

Una buena explicación sobre el significado de las palabras debe poder dar cuenta de la diferencia entre cómo se codifica la homonimia y cómo se codifica la polisemia. En este trabajo estudio tres teorías sobre la polisemia: la explicación tradicional o de la selección de sentidos (SEL), la explicación del léxico generativo (LG) y la de la teoría de la relevancia sobre los conceptos (TR).

Selección de sentidos (SEL)

SEL afirma que los distintos sentidos de una palabra polisémica están representados en paralelo en el léxico. Hay una representación para cada significado de la palabra (Katz 1972). En el caso de (1) tendríamos algo parecido a lo siguiente:

$Banco_1$: lugar en el que sentarse.

$Banco_2$: edificio en el que se realizan operaciones financieras

$Banco_3$: empresa o entidad financiera

Estos tres significados estarían codificados en paralelo en el léxico en una entrada léxica distinta.

Algunos autores han observado varios problemas con SEL. Primero, no puede dar cuenta del uso creativo de las palabras (Pustejovsky 1995), ya que todos los significados tienen que estar almacenados previamente a su uso. Otro problema es que el modelo exige al hablante una gran capacidad de almacenaje (Vicente y Falkum 2015). Tendría que estar codificados en el léxico cada significado de cada una de las palabras que hay. Además, SEL no puede explicar los resultados empíricos antes citados (Frisson 2009; Klepostiniotou et al. 2008; Pylkkänen 2006), porque si los distintos significados se codifican en entradas separadas, entonces, no habría diferencia entre cómo se codifican los significados homónimos y polisémicos.

Léxico generativo (LG)

LG es un sistema de cuatro niveles: la estructura argumental (ARG), la estructura de eventos (EV), la estructura qualia (QU) y la estructura de herencia léxica (HL). La semántica de un ítem léxico es una estructura formada por los cuatro componentes conectados por un conjunto de dispositivos selectivos. La polisemia típicamente se forma a través de mecanismos generativos que ocurren dentro del léxico a

partir del significado más común del término a otros sentidos que se superponen. Sea α un ítem léxico cualquiera:

(2)

$$\begin{bmatrix} \alpha \\ EV = & \begin{bmatrix} E1 = e1 \\ Head = e1 \end{bmatrix} \\ & \dots \\ ARG = & \begin{bmatrix} ARG1 = x \end{bmatrix} \\ & \dots \\ QU = & \begin{bmatrix} Constitutivo: material\ de\ x \\ Formal: qué\ es\ x \\ Télico: función\ de\ x \\ Agentivo: origen\ de\ x \end{bmatrix} \\ \dots \end{bmatrix}$$

Un ejemplo ilustrativo es el de la polisemia del verbo "bake":

(3)
(3a) Bake a cake
(3b) Bake a potato

Los significados de "bake" en (3) son diferentes. En (3a) "bake" contiene el evento de creación, ya que la tarta no es tarta hasta que no se hace (o se hornea). En (3b) "bake" indica un cambio de estado de la patata de cruda a asada. El significado menos específico de "bake" es el de cambio de estado, que es representado según el análisis de LG de la siguiente forma:

(4)

$$\begin{bmatrix} \textbf{bake} \\ EV = & \begin{bmatrix} E1 = e1: proceso \\ HEAD = e1 \end{bmatrix} \\ ARG = & \begin{bmatrix} ARG1 = & \begin{bmatrix} iniciador - animado \\ FORMAL = objeto - físico \end{bmatrix} \\ ARG2 = & \begin{bmatrix} MASA \\ Formal = objeto - físico \end{bmatrix} \end{bmatrix} \\ QU = & \begin{bmatrix} cambio - de - estado \\ AGENTIVO = acto - bake(e1, ARG1\ ARG2) \end{bmatrix} \end{bmatrix}$$

La estructura de "cake" es la que sigue:

(5)

$$\begin{bmatrix} \textbf{cake} \\ ARG = & \begin{bmatrix} x: comida \\ y: masa \end{bmatrix} \\ QU = & \begin{bmatrix} Constitutivo = y \\ Formal = x \\ Télico = COMER\ (e1, z, y) \\ Agentivo = acto - bake\ (e1, w, y) \end{bmatrix} \end{bmatrix}$$

El proceso de co-composición entre (4) y (5) genera el QU de "creación" de (6), lo que da lugar al nuevo significado de "bake" en bake a cake:

(6)

$$\begin{bmatrix} \textbf{bake a cake} \\ EV = & \begin{bmatrix} E1 = & e1: & proceso \\ E2 = & e2: & estado \\ HEAD = & e1 \end{bmatrix} \\ ARG = & \begin{bmatrix} ARG1 = & \begin{bmatrix} iniciador - animado \\ Formal = objeto - físico \end{bmatrix} \\ ARG2 = & \begin{bmatrix} artefacto \\ CONST = ARG3 \\ FORMAL = objeto - físico \end{bmatrix} \\ ARG3 = & \begin{bmatrix} MATERIAL \\ Formal = masa \end{bmatrix} \end{bmatrix} \\ QU = & \begin{bmatrix} CREACIÓN \\ formal = existencia\ (e2, ARG2) \\ AGENTIVO = acto - bake(e1, ARG1, ARG3) \end{bmatrix} \end{bmatrix}$$

El sistema de Pustejovsky tiene algunos problemas. Falkum (2007) observa que LG no puede explicar que haya lecturas optativas, ya que según LG el proceso está guiado por leyes generativas internas y obligatorias. Aunque su sistema explica (3a) y (3b), no podría explicar (3c):

(3c): Bake a pizza

La interpretación de "bake" en (3c) como cambio de estado o como creación es optativa, ya que podemos estar hablando de una pizza que hacemos en casa (la creamos) o de una congelada (cambiamos su estado). Pustejovsky (1998) responde a esta crítica que el mecanismo no tiene que ser obligatorio. Sin embargo, resulta difícil entender cómo puede no serlo si el proceso es enteramente interno al léxico. Para que fuese opcional, la generación del sentido tendría que ser sensible al contexto extralingüístico.

LG falla al dar cuenta de la diferencia entre información léxica y conocimiento del mundo (Fodor y Lepore 1998; Falkum 2007; 2011). Si decimos "bake a trolley", la interpretación natural no es la de creación, aunque "trolley" sea un artefacto. Nuestro conocimiento del mundo tiene un papel importante en este proceso.

LG no puede explicar otros fenómenos, como que ciertos usos de un verbo permitan la alternancia causativa y otros no. Así, por ejemplo, Pustejovsky puede explicar que "romper" entra en alternancia causativa (7) y en (8) no:

(7)
(7a) Rompí la cuerda.
(7b) La cuerda se rompió sola.

(8)
(8a) Juan rompió la ley.
(8b) La ley se rompió sola.

La posible propuesta de LG es que el QU de cuerda es de objeto material y el de ley es abstracto, lo que provoca un cambio en el significado de romper. Sin embargo, esto no explica (9)-(10):

Polisemia e infraespecificación semántica

(9a) María rompió el libro
(9b) El libro se rompió solo

(10)
(10a) Aquel suceso rompió nuestra amistad
(10b) Con el tiempo, nuestra amistad se rompió

Aunque libro sea un objeto físico, no diríamos (9b), pero sí (9a). La diferencia se debe a nuestro conocimiento del mundo. Sabemos, por ejemplo, que una cuerda puede romperse por sí sola debido al paso del tiempo y el desgaste, pero aunque podemos decir que el libro se desgasta por el tiempo, no diríamos que se rompe solo. Por otro lado, aunque "amistad" no designe un objeto físico, (10a) y (10b) tienen ambas sentido. Estos casos sugieren que la información que Pustejovsky (1995) considera en su análisis semántico de la palabra no es suficiente. El concepto complejo codificado, además de los distintos sentidos de las palabras, debe contener información sobre nuestro conocimiento del mundo, que el análisis de LG no incluye.

Teoría de la relevancia (TR)
Otra explicación es TR (Wilson 1998; Carston 2002; Carston y Wilson 2006) donde los sentidos polisémicos de una palabra se forman a partir de mecanismos pragmáticos que dan lugar a conceptos ad hoc atómicos.

La tesis de TR es que la interpretación léxica envuelve la construcción de un concepto ad hoc, o sentido específico de la ocasión, basado en la interacción entre conceptos codificados, información contextual y expectativas o principios pragmáticos. El concepto ad hoc se puede formar estrechando o ampliando el significado lingüísticamente especificado. Los significados polisémicos son también significados específicos de la ocasión o conceptos ad hoc. De tal forma que los significados polisémicos de "banco" -$banco_2$ y $banco_3$- son conceptos ad hoc que surgen a partir de la interacción entre otros conceptos codificados, información contextual y expectativas pragmáticas. En algunos casos estos conceptos se convencionalizan, representándose el concepto ad-hoc y el concepto codificado como entradas léxicas distintas con la misma forma lingüística. En otros, habrá un solo significado codificado de la palabra y sus diferentes usos en distintos contextos se deben a ajustes pragmáticos del significado previamente codificado. Esta tesis distingue dos clases de polisemia. Por un lado, la polisemia pragmática, en la que los distintos usos polisémicos de la palabra se deben a mecanismos pragmáticos del significado previamente codificado. Por otro lado, la polisemia semántica, en la que los dos conceptos se representan en dos entradas distintas. "Banco" tendría que ser un caso de polisemia semántica, ya que ambos significados de la palabra (edificio en el que se saca dinero y empresa o entidad financiera) están convencionalizados. De modo que los dos significados $banco_2$ y $banco_3$ se codifican por separado en conceptos atómicos distintos, igual que en $banco_1$ –que es homónimo con $banco_2$ y $banco_3$-.

El problema que plantea esta hipótesis es que en el caso de la polisemia pragmática, no se puede distinguir una palabra polisémica de la modulación contextual. El significado polisémico sería un significado creativo de la palabra y dependiente del contexto. Por otro lado, en el caso de la polisemia semántica, los significados polisémicos se codifican igual que ocurre en la homonimia. Se crea un concepto atómico separado del concepto previamente codificado. Los significados de 'banco' –$banco_1$, $banco_2$ y $banco_3$- se codificarían todos de la misma manera. Esta hipótesis no puede dar cuenta de los resultados empíricos de los estudios (Frisson 2009; Kleposniotou et al. 2008; Pylkkänen 2006), según los cuales, mientras que los significados homóminos se codifican en paralelo, los significados polisémicos no pueden codificarse de esta forma. Este problema se debe a que la TR propone que las palabras codifican conceptos atómicos, de modo que los distintos significados se tienen que codificar en conceptos separados unos de otros. En artículos posteriores Carston (2016) propone una tesis alternativa, en la que las palabras no codifican conceptos sino un significado infraespecífico y en la que los conceptos ad hoc aparentan ser los auténticos conceptos. Sin embargo, la tesis aún no es definitiva y sigue presentando algunos problemas. No queda claro que Carston pueda distinguir entre polisemia semántica y pragmática como ella pretende hacer. Mi propuesta para solucionar los problemas que se le plantean a la teoría de la relevancia es que los significados no codifican conceptos atómicos, sino complejos.

Conceptos complejos
Para solucionar los problemas de TR y de LG, sugiero que las palabras no codifican conceptos atómicos, sino compuestos. Los conceptos compuestos tienen que contener los distintos sentidos o rasgos de los significados de las palabras.

Los significados polisémicos y homónimos estarían codificados de distinta forma: mientras que los significados homónimos se codificarían en entradas léxicas distintas, los significados polisémicos estarían codificados dentro de una misma entrada léxica. En esta misma línea, Vicente y Martínez-Manrique (2014) sugieren que los distintos sentidos polisémicos están unificados bajo una sola representación conceptual.

La propuesta de este trabajo es que los significados polisémicos se codifican bajo una misma entrada léxica, conformando un concepto complejo en el que hay codificada información de distinto tipo. Por ejemplo, "banco" codificaría, al menos, dos conceptos complejos: $banco_1$ y $banco_2$ correspondientes (1a) con $banco_1$ y (1b)-(1c) con $banco_2$. $Banco_2$ debe incluir no sólo los distintos polisémicos, sino toda la información relevante. Algunos de estos rasgos serán comunes entre (1b) y (1c). La diferencia entre la homonimia y la polisemia es que mientras que los dos significados homónimos que caen bajo la forma ´banco´ se codifican en conceptos distintos; en el caso de la polisemia, los distintos sentidos se codifican en un mismo concepto complejo.

Polisemia e infraespecificación semántica

La solución funciona también con otros ejemplos como ´escuela´, que puede entenderse como el edificio en el que se da clase; como el conjunto de profesores que dan clase en un colegio; como una forma de enseñanza; como una institución; etc. Todos estos sentidos son sentidos polisémicos de una misma palabra y se codificarían en una misma entrada léxica. Toda la información relevante estaría almacenada en un mismo concepto complejo.

La propuesta soluciona el problema que tiene LG con (3c), porque la activación del significado correcto no se genera exclusivamente a partir de la información que Pustejovsky (1995) incluye en el léxico, sino que hay más información involucrada. En el caso de (3) los significados de "bake" se generan a partir de un significado infraespecificado y la información asociada al argumento del verbo. Eso mismo ocurre en (7)-(10), donde los distintos significados de "romper (se)" se deben a las diferencias del argumento del verbo. La alternancia causativa de "romper" funciona dependiendo del significado de su argumento. En (7) hay alternancia causativa debido a la información contenida en "cuerda" y en (8) no la hay debido a la información contenida en "ley". La información relevante contenida en el argumento bno puede limitarse a la que se muestra en (2). La diferencia con LG es que hay más información relevante de la que se considera en (3) que forma parte del concepto complejo. Que esta información es relevante para explicar fenómenos semánticos se ve en otros casos que afectan a verbos y que son similares a "bake" en que la información asociada al argumento afecta al verbo, como lo es en los ejemplos (7)-(10). Estos casos sugieren que la información que Pustejovsky (1995) considera en su análisis semántico de la palabra no es suficiente.

El concepto complejo codificado, además de los distintos sentidos de las palabras, debe contener información sobre nuestro conocimiento del mundo, que el análisis de LG no incluye. La información contenida sobre el argumento del verbo no puede limitarse a si se trata de un objeto físico, abstracto, su origen, etc. Hay mucha más información relevante para la formación del nuevo sentido "romper". El sentido de romper se generaría a través de mecanismos generativos a partir de esa información almacenada en el concepto complejo del argumento del verbo. Esta información almacenada tiene que ser suficiente para poder causar alternancia causativa en (10) pero no en (9). El concepto del argumento del verbo tiene que ser un concepto complejo, que almacene datos sobre qué sabemos de la amistad, las cuerdas, las leyes y los libros y el conocimiento que tenemos sobre cómo pueden o no romperse esos objetos.

Mi hipótesis es que los significados polisémicos se codifican como partes o rasgos de un concepto complejo que contiene toda la información relevante sobre ese significado. Los significados homónimos de la palabra se codificarían en un concepto distinto, como si se tratase de otra palabra. De esta manera, al interpretar sustantivos como "banco" o "escuela" se seleccionarían los rasgos relevantes del significado almacenados en ese concepto. La polisemia de los verbos parece, en cambio, depender de la polisemia del argumento del verbo, como es en el caso de "bake", donde los distintos sentidos se originan a partir de mecanismos generativos que parecen depender de la información contenida en el concepto del argumento del verbo.

Agradecimientos: Quiero agradecer a Agustín Vicente sus aportaciones y discusiones que han llevado a elaborar este trabajo y a Cristina Corredor por sus comentarios en la presentación del Congreso Graduados 2016. El contenido de estas páginas encuadra dentro del grupo consolidado Hizkuntzalaritza Teorikoko Taldea (HiTT, Ref. IT769-13) y en el proyecto de investigación FFI2014-52196-P de MINECO.

Referencias

Katz, J. J. (1972): *Semantic Theory*, New York: Herper and Row.
Carston, R. (2002): *Thoughts and Utterances*. London: Blackwell.
Carston, R. (2016): 'The heterogeneity of procedural meaning'. *Lingua*, 175, 154-166.
Wilson, D., y Carston, R. (2006): 'Metaphor, relevance and the 'emergent property' issue'. *Mind and Language*, 21(3), 404-433.
Falkum, I. L. (2007): 'Generativity, relevance and the problem of polysemy'. En: UCL *Working Papers in Linguistics* Vol. 19, pp. 205-234.
Falkum, I. L. (2011): *The Semantics and Pragmatics of Polysemy: A Relevance-Theoretic Account*, University College London.
Fodor, J. A., and Lepore, E. (1998): 'The emptiness of the lexicon: rejections on James Pustejovsky's The Generative Lexicon', *Linguistic Inquiry* : 29(2), 269-288.
Frisson, S. (2009): 'Semantic underspecification in language processing'. Language *and Linguistics Compass*, 3(1), 111-127.
Klepousniotou, E., Titone, D. y Romero, C. (2008): 'Making sense of word senses: the comprehension of polysemy depends on sense overlap'. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 34(6), 15-34.
Pylkkänen, L., Llinás, R., y Murphy, G. L. (2006): 'The representation of polysemy: MEG evidence'. *Journal of cognitive neuroscience*, 18(1), 97-109.
Pustejovsky (1995): *The Generative Lexicon*. Massachusetts: THE MIT PRESS.
Pustejovsky (1998): 'Generativity and explanation in semantics: A Reply to Fodor and Lepore', *Linguistic Inquiry*, 29(2), 289:311
Vicente, A y Falkum, I. L. (2015): 'Polysemy: Current perspectives and approaches', *Lingua*: 157, 1-16.
Vicente, A., y Martínez-Manrique, F. (2014): 'The big concepts paper: a defence of hybridism'. The *British Journal for the Philosophy of Science*, 67(1)
Wilson, D. (1998): 'Linguistic structure and inferential communication'. En: *Proceedings of the 16th International Congress of Linguists* (pp. 20-25). Oxford: Elsevier Science

III Conferencia de Graduados de la SLMFCE

How to fix what is said

Claudia Picazo Jaque
Universidad de Barcelona (LOGOS)

Abstract: Contextualists have shown that, in general, the content of an utterance of a sentence S need not be equivalent to the linguistic meaning of S. The question I will address here concerns the determination of the truth-conditions of an utterance when these are underdetermined by linguistic meaning. First, I will argue that they are not determined by the speaker's intentions. Second, I will argue that, if it is assumed that they are determined by an objective parameter, then it needs to be explained how the value of that parameter is determined. Third, I will put forward a view in which the truth-conditions of the utterance partially depend on how it is reasonable to interpret the utterance.

Keywords: contextualism, what is said, speaker's intentions, interpretation.

It is nowadays common to hold the view that the content of an utterance of a sentence S need not be (or cannot be, according to Radical Contextualism) equivalent to the linguistic meaning of S, even in absence of indexicals. Thus, the sentence 'Tipper is ready' can say that Tipper is ready for dinner or that Tipper is ready for an interview; 'Hugo weighs 79 kilos' can say that Hugo weighs 79 kilos naked, before breakfast or that Hugo weighs 79 kilos with his clothes on; 'Paul needs a red pen' can say that Paul needs a pen with red ink or that Paul needs a superficially red pen, and so on. The outcome is that the property expressed by the predicate in these cases is underdetermined by linguistic meaning: the linguistic meaning of 'is ready', 'weighs 79 kilos' and 'needs a red pen' is compatible with expressing a variety of properties. In order to avoid some complications concerning properties, we can talk about shifts in the application or satisfaction conditions of predicates (or in what counts as 'is ready', etc.).

As a result of these shifts, we need to distinguish three notions: (i) linguistic meaning, (ii) what is said, (iii) implicatures. The distinction raises important questions, including the following: How do we individuate what is said from implicatures? Is what is said identical to what is asserted? Should what is said be construed as a variety of speaker meaning, as implicatures? How is what is said by a sentence in an occasion of use determined? In this talk, I will focus on the latter. However, before that, it is important to note that the relevance of the notion of what is said goes beyond the debate about the semantics/pragmatics divide. For instance, this notion plays a role in the epistemology of testimony (since what is asserted plausibly goes beyond linguistic meaning), it grounds joint action (in ordinary situations, we coordinate with each other by means of enriched propositions) and it often is what speakers are liable for (we hold speakers responsible not only for the linguistic meaning of their utterances, but also for what their utterances, in context, say). Thus, the notion of what is said plays a crucial role in any explanation of our communicative practices.

The question I will address here concerns the determination of the application conditions of predicates as the ones mentioned in the previous examples. My aim is to argue that the content of an utterance of a sentence (what is said), when underdetermined by semantics, is (i) not determined by the speaker's intentions or some interpreter-independent property of the context of use but instead (ii) it depends how it is reasonable, given the available information, to interpret the utterance. In what follows, I will assume that the content of the utterance in question is not determined by semantics, without addressing the question whether this is so because semantics is unfit to provide a truth-evaluable content or because the intuitive truth-conditions of the utterance go beyond the content semantically determined. This assumption is justified by the amount of examples contextualists have provided over the last decades.

The accounts one can find in the literature concerning how is the content of an utterance determined can be classified into three kinds. First, some philosophers (Bach 1994, Carston 2002, Borg 2012, Perry 2009) take the content of an utterance to be either determined by or identical to what the speaker intends to communicate. I will call such accounts Speaker Centred Views. According to these views, the truth-conditional content of an utterance of 'Tipper is ready', etc., is contingent upon what the speaker intends to communicate. The content of the utterance is constrained by its linguistic meaning, but is modulated in accordance with the speaker's communicative intentions. If by uttering 'Tipper is ready' the speaker meant that Tipper is ready for an interview, then that is the content of his utterance.

Second, some philosophers take the content of an utterance to be determined by linguistic meaning together with some objective parameter of the context of use. I will call these approaches Interlocutor Independent Views. It is common to talk about purposes. Here, I will focus on Stokke and Schoubye's (2015) account, since it is, to my knowledge, the only one that directly addresses the question about how to determine what is said. Schoubye and Stokke present a theory according to which what is said is determined by the semantic meaning of the sentence, together with a question under discussion (roughly, the topic of the conversation). Contexts, in Schoubye and Stokke's view, contain questions under discussion. The goal of the conversation is to answer them. What is said by a sentence S in a context c relative to question q_c is the weakest relevant proposition p such that p entails the semantic content of S in c. Thus, if the question under discussion in a given context is 'Is Tipper ready for the interview?' the content of

How to fix what is said

an utterance of 'Tipper is ready' will be that Tipper is ready for the interview.
Third, some philosophers have put forward what I will call Interpreter-Oriented Views. I include here Recanati's (2004) availability principle and views that identify content with reasonable interpretation (Travis 1989, Gauker 2008 for demonstratives). According to these views, the content of an utterance is fixed by how it is plausible or reasonable to interpret it. These views do not provide a mechanism that decides, given a context and a sentence, what would be the content of an utterance of that sentence in that context. Instead, they claim that we have to decide the issue case by case, for we cannot systematise all the information that might be relevant for interpreting an utte rance.I will put forward a mixed view in which interpreters' judgements play a crucial role in the determination of what I will call the activity in place but in which, once this is determined, the determination of the satisfaction conditions of a predicate no longer depend on the interlocutors.

Contrary to Speaker Centred Views, there is no reason to assume that speakers have any special authority about the content of their utterances (what their utterances say), although they have, of course, authority about what they mean (what they intend to say by an utterance). SCV construes 'what is said' as a notion closely tied to speaker's intentions. However, if we focus instead on the roles the notion of what is said is supposed to play, we see that we need rather a socio-linguistic notion. I will focus on the way we attribute communicative responsibilities. When a speaker engages in a conversation, and asserts something, he is thereby putting forward some information that the hearer might use in order to rationally choose among different courses of action. We use what others tell us in order to plan our behaviour, and hold them responsible when the information they provide is inaccurate. This holds for other speech acts, such as requests, promises...

A simple example. María's pen has run out of ink, and she says to Luis: 'I need to write some corrections on this report in green, so I'm going to buy a new green pen. Do you need something?' Luis replies: 'I need a red pen. Could you buy one for me? I will pay you back later'. In this example, the salient interpretation of 'red pen' is pen with red ink. If María buys such a pen, then, because of the previous conversation, Luis should pay for it, regardless of what he had in mind when he spoke. We can imagine that he in fact meant to ask for a superficially red pen. However, he has acquired a certain responsibility that goes together with the salient interpretation. This provides evidence against Speaker Centred Views because of two reasons. (i) The simplest explanation of why Luis should pay the money is that María has fulfilled his request. (ii) Luis could insist that he didn't ask for that kind of pen, but the most natural way to do that would be to insist that he merely asked for a 'red' pen, not that what he really asked for was a superficially red pen (thus, liability goes either with linguistic meaning or with the most salient or available interpretation). In view of this phenomenon, Perry (2009) has argued that in semantics we need to substitute what he calls the forensic notion of what is said by an intention-based notion. Against this, one important aim in doing semantics is to understand our linguistic practices. Semantics needs therefore to connect with use, and if we base semantics in a notion of content that does not correspond with our ordinary notion, we loose the connection. Moreover, we already have an intention-based notion: what the speaker means.

On the other hand, the problems concerning Interlocutor Independent Views are (i) explaining how is the alleged parameter determined and (ii) whether there really is one and only one parameter determining the content of our utterances (beyond semantics). Unless one explains how are, for instance, purposes, determined, one has not provided a framework that accounts for the determination of what is said: one has only given half the answer. Moreover, if we go back to questions under discussion, Stokke and Schoubye have not managed to convincingly argue for the claim that the content of an utterance is determined by a question under discussion. Against their claim, there are cases in which there is an explicit question being discussed but where, because of other available information, it is more natural to interpret an utterance independently of that question. For example, cases in which it is clear to everybody in the conversation that the speaker is not answering the question under discussion but is talking about something else (because of a gesture, let's say). The utterance, then, can be said to change the context, but if so, the new context does not contain a question under discussion: the utterance in need of interpretation would itself be providing the topic of the new conversation.

The examples used in the previous arguments motivate two claims. First, given that we attribute responsibilities and obtain information from the content of other speaker's utterances, whatever determines content must be available to us. Otherwise, we would disconnect the semantic notion of content from these practices. The metaphysics of content cannot outrun its epistemology, so to speak. This claim is independently motivated by the Davidsonian principle that language being essentially public, meaning cannot outrun our interpretative capacities. Second, a variety of things might matter to the identification of what is said: the topic of the conversation, previous discourse, general knowledge about the world, gestures, salient objects in the situation in which the conversation takes place, etc. It is very unlikely that we can establish a mechanism determining how these elements will matter in a possible conversation.

However, as Gauker has argued for the case of demonstratives, we can have a general principle establishing how is the content of an utterance fixed: the content of an utterance is given by an all-things-considered judgement, where the things to be considered include all the available infor-

III Conferencia de Graduados de la SLMFCE

mation about the context of use (to a hearer paying full attention). Thus, what is said is can be identified with what is available to normal speakers (Recanati) or with the interpretation of a reasonable speaker (Gauker, Travis), under ideal conditions.

This option has some advantages, but is not completely satisfactory. It clashes with the intuition that, in a conversation about going out for dinner, what makes it the case that an utterance of 'Tipper is ready' is true if and only if Tipper is ready for dinner is the topic of the conversation. Reasonable or normal speakers will arrive at this interpretation, because content does not outrun our capacities, but the determination of content need not go via our interpretations.

A mixed model can work better. Given that it seems unlikely that there is an objective mechanism determining what is said by an utterance, we need to allow some role for reasonable interpreters. But their role will be the identification of the element that determines what is said. My proposal goes as follows: utterances take place within broader activities (planning to go out for dinner, writing reports…). Which activity is in place depends on an indefinite number of features of the context we can't list. We can let interlocutors' judgements decide which is the activity in place: the activity in place is the most reasonable one, given the available information. Once the activity has been identified, the satisfaction conditions of a predicate are determined by it. For example, in talking about going out for dinner, only people who are dressed, etc. count as ready; in writing reports, only pens with red ink count as red, etc. Given that activities are something we easily recognise and, as a rule, know what matters in them, what is said is available to normal interpreters. Moreover, the notion thus obtained is in line with linguistic liability.

References:

Bach, Kent (1994). Conversational Impliciture. *Mind and Language* 9 (2):124-162.
Borg, Emma (2012). *Pursuing Meaning*. OUP Oxford.
Carston, Robyn (2002). *Thoughts and Utterances*. Blackwell.
Gauker, Christopher (2008). Zero tolerance for pragmatics. *Synthese* 165 (3):359–371.
Perry, John (2009). Directing intentions. In Joseph Almog & Paolo Leonardi (eds.), *The Philosophy of David Kaplan*. Oxford University Press 187--201.
Recanati, François (2004). *Literal Meaning*. Cambridge University Press.
Schoubye, Anders J. & Stokke, Andreas (2015). What is Said? *Noûs* 49 (4).
Travis, Charles (1989). *The Uses of Sense: Wittgenstein's Philosophy of Language*. Oxford University Press.

SOCIEDAD DE LÓGICA, METODOLOGÍA Y FILOSOFÍA DE LA CIENCIA EN ESPAÑA

Octubre de 2016

www.solofici.org

www.solofici.org