

EPISTEMOLOGÍA E HISTORIA DE LA CIENCIA

SELECCIÓN DE TRABAJOS DE LAS XI JORNADAS

VOLUMEN 7 (2001), Nº 7

Ricardo Caracciolo

Diego Letzen

Editores



ÁREA LOGICO-EPISTEMOLÓGICA DE LA ESCUELA DE FILOSOFÍA
CENTRO DE INVESTIGACIONES DE LA FACULTAD DE FILOSOFÍA Y HUMANIDADES
UNIVERSIDAD NACIONAL DE CÓRDOBA



Esta obra está bajo una Licencia Creative Commons atribución NoComercial-SinDerivadas 2.5 Argentina



Integración de niveles de coherencia en la comparación de teorías y aproximaciones científicas

Gabriel Painceyra / Hernán Severgnini*

Introducción

Dentro de la dinámica histórica del descubrimiento científico se hallan involucrados distintos aspectos relacionados con el cambio de las concepciones científicas. Estos cambios constituyen la dinámica de teorías e implican la necesidad de los científicos de “articular” los nuevos descubrimientos dentro de las concepciones o tradiciones de investigación a las que pertenecen. Intervienen reglas heurísticas no sólo durante el proceso de descubrimiento sino también justamente para articular lo descubierto: en este sentido, las heurísticas constituyen procesos inferenciales de diferentes tipos que le permiten al científico la formulación de lo descubierto, por una parte, y su vinculación con conocimientos previos, por otra. Así, si el descubrimiento científico, ligado de algún modo a reglas de búsqueda, se manifiesta finalmente en una dinámica o cambio científico, la comprensión de los momentos que componen dicho cambio también constituye un tema de indagación para quien estudia este tipo de descubrimiento.

Algunos modelos computacionales han tratado de explicitar los criterios de comparación presentes en la dinámica de las teorías científicas. La tesis de la inconmensurabilidad de Kuhn intentaba expresar la relación que surge entre las distintas tradiciones de investigación durante los denominados “descubrimientos revolucionarios”. Esta relación de inconmensurabilidad también ha sido objeto de estudio en el campo de la modelización computacional. El análisis de la reconstrucción ofrecida por Paul Thagard en *Conceptual Revolutions*,¹ que consiste principalmente de su Teoría de la Coherencia Explicativa (TEC) y su implementación computacional (ECHO), arrojó como resultado ciertos alcances y límites.

Entre estos límites hemos señalado ya² que la comparación es imposible de realizar cuando las teorías difieren en aspectos conceptuales y experimentales, o a nivel metodológico general. Las tradiciones de investigación que sostienen distintas concepciones acerca de qué constituye una explicación, de qué es un dato relevante, o de qué es científico y qué no lo es, no pueden ser contempladas por ECHO ya que la emulación computacional no puede dar entrada a este tipo de consideraciones. Esto se corresponde con la distinción que efectúa Thagard entre aproximaciones científicas, marcos teóricos y teorías. Mientras que las teorías son consideradas como meros conjuntos de hipótesis explicativas, las aproximaciones (*approaches*) son mucho más generales, comprendiendo una “colección de métodos experimentales y estilos de explicación.”³ A su vez, una aproximación puede englobar varios marcos (*frameworks*) diferentes, como por ejemplo la aproximación cognitivista en psicología actual, que incluye tanto al marco cognitivista clásico de Simon como también al conexionismo. Por nuestra parte creemos, a pesar de la opinión de Thagard, que la simulación computacional en Inteligencia Artificial no debería limitarse solamente a los casos de comparación de teorías, sino que puede extenderse a los de comparación de marcos y aproximaciones.

* Universidad Nacional de Córdoba.

Otro de los límites de ECHO está en la consideración de la coherencia explicativa teniendo en cuenta sólo los aspectos estructurales de las teorías en competencia. Si bien Thagard comienza su libro exponiendo una noción de "concepto" cercana a la de "frame" de Minsky, no hay ninguna implementación computacional de esta dimensión semántica de las teorías. La reconstrucción computacional es, así, meramente estructural, tomando como unidad mínima de análisis a la proposición entendida sólo sintácticamente, y no a los conceptos con su aspecto semántico.

Posteriormente trabajos de Thagard y sus colaboradores han presentado una serie de elementos y tratamientos de las diferentes clases de representaciones y de las relaciones entre ellas, como así también de la coherencia en sus múltiples manifestaciones, siempre dentro del marco teórico de satisfacción de constreñimientos (*constraints*). Estos nuevos desarrollos permitirían solucionar, al menos parcialmente, algunos de los problemas y límites a los que hemos hecho referencia.

Encontramos entonces dos problemas centrales al evaluar los resultados alcanzados por estos modelos computacionales: por un lado, y ligado a la cuestión de la comparación entre aproximaciones, el problema de cómo integrar diferentes niveles de coherencia (como por ejemplo, coherencia explicativa, analógica, deductiva, perceptual, conceptual) y, por otra parte, el de cómo integrar distintas formas de representación. Ambos problemas están relacionados ya que, en algunos casos, integrar niveles de coherencia requiere previamente integrar diferentes formas de representación.

En este trabajo intentaremos aportar algunas soluciones con respecto a ciertos tópicos de esta problemática. Para ello comenzaremos mostrando y analizando algunos de estos nuevos desarrollos efectuados por Thagard. En particular, mostraremos en qué consiste el llamado "problema de la intercoherencia" haciendo especial mención de su importancia epistemológica y elaborando por nuestra parte un modelo teórico de solución. A este esbozo de solución llegaremos durante el desarrollo de un nuevo tipo de coherencia que denominaremos *Coherencia Evidencial*, ideado con el objeto de extender la simulación a algunos casos de comparación de aproximaciones científicas. A continuación, encararemos el problema de cómo integrar diferentes tipos de representaciones. Aquí, nos limitaremos a realizar una evaluación de una nueva herramienta conexionista de reciente desarrollo, el algoritmo DRAMA, ofrecido como un intento de integración computacional de los aspectos sintácticos y semánticos presentes en la fabricación de analogías.

El problema de cómo integrar diferentes tipos de coherencia, y una solución al problema de la evidencia empírica

Durante estos últimos años Thagard aplicó su teoría de satisfacción de *constraints* a otros dominios, de la misma forma en que lo había realizado en el plano epistemológico con la Coherencia Explicativa para la comparación de teorías. Así ha obtenido modelos para la coherencia deductiva, analógica, perceptual, conceptual, y otras. Sin embargo se presenta ahora el problema de cómo integrar todos estos niveles de coherencia bajo un marco unificador. A este problema lo ha denominado el "problema de la intercoherencia", y es particularmente relevante para nosotros ya que una solución al mismo podría significar también un principio de diseño de nuevos modelos capaces de comparar no sólo teorías proposicionales sino también marcos y aproximaciones científicos. Esto es así ya que podemos considerar a estos niveles más generales de inconmensurabilidad como compuestos de diferentes

tipos de coherencia (como los mencionados arriba), de tal manera que una adecuada modelización de sus interconexiones serviría de base para la simulación de los procesos de comparación de tales marcos y aproximaciones.

Así, en un artículo titulado "Knowledge and coherence",⁴ Thagard señala que es necesario el diseño de una teoría multifacética de la coherencia, y aclara además que la dificultad para ésta reside más específicamente en cómo integrar diferentes tipos de coherencia que se evalúan sobre la base de representaciones de distinta clase. El caso concreto es que, si bien es factible evaluar conjunta y simultáneamente coherencias de tipo explicativa, deductiva y analógica (aunque Thagard no ofrece ningún modelo ni solución al respecto), dado que comparten la característica de estar expresadas a través de conjuntos de proposiciones, no está claro, en absoluto, cómo integrarlas con los tipos de coherencia que utilizan elementos no proposicionales, como los visuales y conceptuales.

Uno de los aspectos más importantes que se presenta en los casos de comparación entre aproximaciones científicas consiste en que en estos casos la evidencia empírica no es compartida. Esto significa que los proponentes de las distintas aproximaciones ni siquiera acuerdan con respecto a cuáles son los datos empíricos que hay que explicar. En situaciones de esta índole, por consiguiente, uno de los principios teóricos clave de TEC, el que daba prioridad computacional a los datos de evidencia, ya no puede sostenerse más (al menos, tal como está implementado en ECHO), puesto que lo que está ahora en cuestión es, justamente, qué es lo que considera como dato empírico cada una de las partes que compiten por la mejor coherencia explicativa. A nuestro parecer, si bien ECHO no puede realizar esta comparación, los científicos la realizan en determinadas ocasiones y con justificadas pretensiones. Por ejemplo, el conductismo no acepta, como evidencia científica que deba ser explicada, a las preferencias verbales de los sujetos observados, a no ser que sólo sean consideradas como conductas motoras sin significado lingüístico. Estamos aquí ante un criterio específico y particular de lo que se considera "evidencia científica". Por otra parte, el cognitivismo toma precisamente, como elementos que componen su evidencia científica, a las preferencias verbales con su significado lingüístico. Puesto que la comparación y elección de una aproximación particular se realiza sobre bases racionales, creemos que, más allá de los límites de ECHO, una simulación de tales procesos debería ser posible.

El esbozo de solución que proponemos para este tópico parte de lo que Thagard mismo presentó en "Knowledge and coherence" como Coherencia Perceptual. Ésta se construye sobre la base de relaciones entre *inputs sensoriales* e *interpretaciones* visuales. Basándose en las leyes de la *Gestalt*, los principios que rigen este tipo de coherencia relacionan ciertas características de los inputs sensoriales, que deben ser descriptos adecuadamente, con el modo como tienden a interpretarse. Además, los inputs sensoriales tienen aceptabilidad por sí mismos, es decir, tienen prioridad computacional, tal como la tenían los datos de evidencia científica en TEC y en ECHO.

Como ya lo expresamos, en el caso de aproximaciones científicas incommensurables no podemos sostener la "prioridad" de los datos de evidencia, dado que gran parte de la evidencia que es aceptada por una aproximación no lo es por parte de la otra, o es interpretada de manera diferente. Pero consideramos que el problema puede ser resuelto si dejamos de lado tal principio de prioridad, que implica considerar a cada pieza de evidencia como exenta de disputa científica, y establecemos, por el contrario, la posibilidad de que su aceptabilidad esté dada por ciertos criterios presentes en los cánones científicos de las co-

comunidades de investigación. De esta manera, una solución natural para los problemas basados en inconmensurabilidad científica de este tipo consistiría en desdoblar el problema de la evidencia, y considerarla como compuesta de *inputs sensoriales* (o, más en general, de *inputs de carácter empírico*), por un lado, e *interpretaciones*, por otro. Así estaremos en condiciones de simular las diferencias existentes entre distintas aproximaciones con respecto a la interpretación de las piezas de evidencia.

Por lo tanto, creemos que el antiguo principio de "prioridad de los datos" de TEC debe analizarse de la siguiente manera: cada evidencia debe tomarse como un par ordenado de interpretaciones teóricas, por un lado, e inputs que, en algunos casos serán directamente inputs sensoriales y, en otros, reportes empíricos, informes experimentales u observaciones de alguna clase. Estos pares ordenados aparecerán entonces como (E1, IE1) y (E'1, IE1), donde E1 y E'1 son elementos de una red conexionista del tipo de ECHO que se corresponden con las distintas interpretaciones en competencia; e IE1 es el elemento de la red que representa al input de carácter sensorial o empírico que se intenta interpretar científicamente dentro de un marco teórico o aproximación científica (ver Figura 1).

A continuación, la teoría de este nuevo tipo de coherencia que denominamos "coherencia evidencial", establecerá los *constraints* que se dan entre los diferentes elementos (interpretaciones e inputs) y que explicitarán los criterios epistémicos de evidencia empírica de las distintas comunidades científicas en competencia, estableciendo así las relaciones positivas y negativas de coherencia evidencial. Partiendo de esta teoría, entonces, el nuevo algoritmo establecerá un link excitatorio entre cada elemento de los pares ordenados, esto es, entre E1 e IE1, y entre E'1 e IE1. De esta forma simulamos el hecho de que cada aproximación científica sostiene interpretaciones propias (E1 y E'1) de cada uno de los inputs empíricos. Estas interpretaciones, obviamente, son incompatibles en algún lugar (expresando así las diferencias e inconmensurabilidades que se dan entre las aproximaciones) y, por lo tanto, en vez de un link excitatorio se establece un link inhibitorio entre los elementos de la red que les corresponden (es decir, entre E1 y E'1). Como elemento central de "prioridad" para ambos marcos se dejan los inputs de carácter empírico que se comparten (en nuestro caso, IE1). Ciertamente no todos los inputs empíricos son relevantes para ambas aproximaciones. Este es un problema de "dominios de aplicación" de cada aproximación. Pero la comparación de todos modos puede realizarse, al menos parcialmente, partiendo de los inputs que sí se comparten.

Esta solución agrega un nuevo nivel de coherencia que es anterior al nivel de la coherencia explicativa. Sin embargo, creemos que esta solución puede ser implementada computacionalmente de dos maneras diferentes. Por un lado podría desarrollarse, en primer término, una evaluación de coherencia *evidencial*, basada en esta distinción entre interpretaciones e inputs de carácter empírico con un algoritmo conexionista denominado "EVIDECHO", y luego una evaluación a nivel proposicional, entre hipótesis explicativas e interpretaciones teóricas de la evidencia considerada científica (ahora sí, a nivel de ECHO, esto es, al nivel de la coherencia explicativa). O, alternativamente, realizar ambas evaluaciones en paralelo y simultáneamente, activando finalmente una intercoherencia entre ambas mediante la estabilización de una sola y misma red conexionista, aunque más compleja que la anterior correspondiente a ECHO. Creemos que esta segunda solución es la más natural desde un punto de vista plenamente conexionista, por más que la construcción de tales redes podría llegar a insumir una cantidad considerable de nodos y conexiones. Ob-

servamos, finalmente, que esta segunda propuesta constituye a su vez un modelo de solución para todos los demás problemas de intercoherencia que preocupan a Thagard.

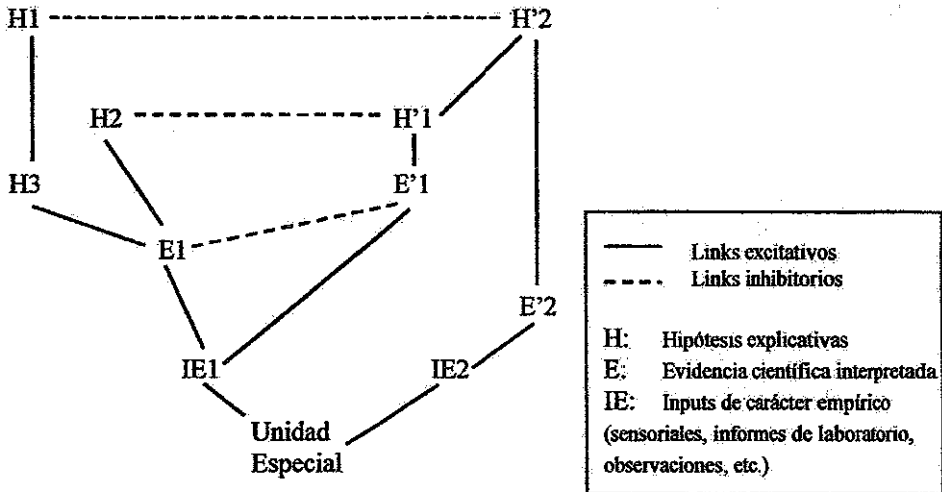


FIGURA 1

Representaciones holográficas reducidas: una alternativa para la integración de estructura y semántica

Hemos mostrado, así, una posible solución al problema de cómo integrar dos tipos diferentes de coherencia: coherencia explicativa y lo que hemos denominado coherencia *evidencial*. El desarrollo de este nuevo tipo de coherencia se nos impuso en el intento de ampliar el dominio de simulación de la teoría, abarcando los casos de comparación entre lo que Thagard denomina “marcos teóricos” y “aproximaciones científicas”. De esta manera, TEC deja de estar limitada a los casos de comparación de teorías, consideradas como meros conjuntos de proposiciones, y se extiende a la elucidación y simulación de los aspectos metodológicos que caracterizan a diferentes comunidades científicas. Por otra parte y al mismo tiempo, hemos forjado una solución teórica a cómo integrar dos formas distintas de coherencia, solución que puede servir de modelo para resolver el denominado “problema de la intercoherencia”.

Sin embargo, nos queda por considerar un problema más básico aún: el de cómo integrar diferentes formas de representación, en particular, cómo integrar representaciones conceptuales y proposicionales, problema que Thagard no podía elucidar en su *Conceptual Revolutions*, y al cual recién ahora parece haber encontrado una respuesta posible a través del algoritmo DRAMA.

DRAMA ha sido desarrollado por Paul Thagard y Chris Eliasmith con el fin de simular la tarea cognitiva de fabricación de analogías.⁵ Este modelo presenta, como aspecto novedoso dentro de la IA, el hecho de que es capaz de considerar tanto aspectos estructurales como semánticos dentro de una misma arquitectura computacional. Esta característica es sumamente importante ya que, como hemos señalado, ECHO, al considerar a las teorías

sólo como conjuntos de proposiciones relacionadas explicativamente, descuidaba completamente el aspecto semántico-conceptual de las teorías.

Sólo con el propósito de mostrar lo que puede constituir una solución posible a este límite de ECHO, expondremos brevemente las características centrales de DRAMA y cómo se da su funcionamiento sobre la base de un nuevo tipo de representaciones distribuidas denominadas "representaciones holográficas reducidas" (HRR).

DRAMA es una red neural que implementa uno de los pasos que han sido identificados en la tarea de fabricación de analogías: el paso denominado en IA como "mapeo analógico" (o, más precisamente, "mapeo de análogos"). El programa tiene como fundamento una teoría computacional que considera múltiples *constraints* al efectuar el mapeo. En particular, la teoría establece tres *constraints* (estructura, similaridad y propósito) que se implementan mediante operadores básicos que operan sobre vectores en un espacio normalizado de 512 dimensiones y que se corresponden, cada uno, con un elemento (concepto, relación o proposición). Los operadores son cuatro:

- 1) suma (composición o superposición de vectores);
- 2) convolución (codificación o producto de vectores);
- 3) correlación (decodificación, función inversa de 2); y
- 4) proyección de un vector sobre otro (o "evaluación de similaridad").

Con estas cuatro operaciones fundamentales DRAMA es capaz de realizar todas las tareas requeridas, esto es, representa proposiciones como combinación de conceptos y también de qué manera se establecen los mapeos entre las situaciones a analogar.

Lo interesante, en relación con ECHO y la Teoría de la Coherencia Explicativa, es considerar el modo en que estas HRR constituyan una alternativa que permita integrar estructura y semántica en la comparación de teorías científicas.

En primer lugar, observemos que las HRR permitirían que las proposiciones de una teoría reciban distintas caracterizaciones según las funciones que cumplen. Por ejemplo, las proposiciones pueden ser convolucionadas gracias a distintas relaciones como "causa", "explica", "subsume" o cualesquiera otras. Esta posibilidad ofrecida por las HRR permite discriminar mejor los distintos tipos de relaciones explicativas, deductivas y analógicas.

Todo el conjunto de proposiciones de una teoría puede ser representado en el mismo espacio normalizado de las HRR, independientemente de si esto es para establecer analogía (que era, en realidad, uno de los principios de TEC para evaluar la coherencia), o para evaluar la coherencia en general. Sin embargo, con HRR las unidades primarias de análisis no son necesariamente las proposiciones sino que pueden serlo también los conceptos. Así, podemos efectuar evaluaciones en paralelo a nivel de la coherencia conceptual y evidencial, o considerar también las distintas relaciones que hacen a la coherencia entre proposiciones partiendo de los conceptos que las componen.

Finalmente, DRAMA permite, por medio de la operación de proyección, la evaluación de la semejanza entre conceptos dentro de una teoría o entre dos teorías o aproximaciones diferentes.

Así, las HRR extienden las posibilidades comparativas de ECHO al representar niveles estructurales y semánticos de las teorías, permitiendo comparar y evaluar niveles de coherencia conceptual, proposicional y explicativa.

En definitiva, creemos que las características de las HRR permiten pensar en soluciones computacionales novedosas y efectivas, e implementarlas dentro del marco teórico del conexiionismo y de satisfacción de *constraints*. Esto es importante si queremos, justamente, ser coherentes con el punto de vista en que han surgido tanto TEC como las demás teorías correspondientes a otras formas de coherencia. Sin embargo, y a pesar de los logros conseguidos, consideramos que aún quedan obstáculos importantes en este camino. En el caso de DRAMA en particular, hemos detectado que, en realidad, el algoritmo será incapaz, al menos en la etapa actual de su desarrollo, de percibir por sí mismo las relaciones estructurales presentes dentro de las proposiciones y aún las relaciones entre proposiciones. La mayor parte de la información sintáctica y estructural sigue siendo alimentada manualmente por el programador mismo. El mérito de DRAMA consiste en ser capaz de codificar, mediante representaciones distribuidas, tal información de carácter sintáctico, y permitir así un substrato computacional común para la manipulación simultánea, y dentro de una misma arquitectura, de aquellos dos niveles tan difíciles de integrar: semántico y estructural. Pero gran parte de los mecanismos cognitivos básicos que involucran al reconocimiento de estructura entre proposiciones y a la sintaxis interna de éstas, aún quedan por ser descubiertos y articulados dentro de una teoría coherente que permita la simulación computacional plena de tales procesos. Asimismo, aspectos tales como el reconocimiento correcto de causas y efectos, o la presencia de una relación explicativa entre dos proposiciones dadas, o incluso la detección misma de una contradicción dentro de un conjunto de proposiciones, son mecanismos cognitivos básicos que no deberían quedar librados siempre al buen juicio del programador de turno, sino formar parte natural de las rutinas de un algoritmo. Y, en este aspecto, tanto TEC como DRAMA, y los demás algoritmos relacionados con ellos, presentan aún una limitación manifiesta.

Desde este punto de vista, entonces, podemos concluir que DRAMA constituye realmente un avance con respecto a ECHO en cuanto a la arquitectura de representación tanto de conceptos como de proposiciones. Aunque no constituye aún un desarrollo más eficaz que su predecesor en cuanto a la manipulación y procesamiento computacional de tales elementos.

Conclusión

Hemos planteado en primer lugar las características y límites de la Teoría de la Coherencia Explicativa y su implementación computacional ECHO para dar cuenta de los procesos cognitivos presentes en la dinámica del conocimiento científico. Estos límites estaban dados por la imposibilidad del modelo de Thagard para ofrecer algún tipo de reconstrucción racional de inconmensurabilidades presentes en los casos de comparación de marcos teóricos o aproximaciones científicas, esto es, los casos más complejos que la mera comparación de teorías científicas.

Por nuestra parte hemos considerado que tales límites se corresponden con lo que Thagard ha denominado posteriormente "el problema de la intercoherencia". A partir de aquí, hemos desarrollado un nuevo tipo de coherencia que llamamos *evidencial*, ofreciendo una solución específica al problema de integrar la coherencia evidencial y la explicativa. Y creemos que esta solución parcial puede generalizarse hacia el caso de la integración de los demás tipos de coherencia que seguramente están presentes en las inconmensurabilidades de tipo metodológico.

Ligado a esto enfrentamos el problema de integrar representaciones de distinto tipo, principalmente conceptual y proposicional, otro de los límites manifiestos en ECHO. En este t3pico consideramos que la soluci3n ofrecida a trav3s del modelo denominado DRAMA, que utiliza HRR, es satisfactoria ya que integra eficazmente la representaci3n tanto de conceptos como de proposiciones, en una arquitectura computacional 3nica. Tambi3n hemos mostrado brevemente c3mo podr3amos aplicar esta herramienta en el caso fundamental que nos ocupa.

Queda por profundizar la investigaci3n sobre la posibilidad de que estos algoritmos sean capaces de reconocer por s3 mismos la estructura sint3ctica de proposiciones, la estructura interna de un conjunto de proposiciones, sus relaciones explicativas, de causa y efecto, de contradicci3n y competici3n, como tambi3n aquellos otros aspectos a tener en cuenta en los casos de comparaci3n de teor3as que presentan inconmensurabilidades locales y metodol3gicas. Y, finalmente, quedan por explicitar tambi3n los principios que rigen la implementaci3n de la Teor3a de la Coherencia Evidencial. Estos ser3n, seguramente, temas de futuros trabajos.⁶

Notas

¹ Thagard, P. (1992), *Conceptual Revolutions*, Princeton, Princeton University Press.

² Cf. Painceyra, G.; y Severgnini, H. (1998), "Sobre una reconstrucci3n computacional de la inconmensurabilidad", en *Epist3me*, vol. 3, n3 7, 139-147.

³ *Op. cit.*, pp. 225-6.

⁴ Thagard, P.; Eliasmith, C.; Rusnock, P., y Shelley, C. P. (en imprenta), "Knowledge and coherence", en R. Elio (ed.), *Common sense, reasoning, and rationality*, Vancouver Studies in Cognitive Science, Vol. 11, New York, Oxford University Press. (Publicaci3n prevista para diciembre 2001, disponible en <http://cogsci.uwaterloo.ca/Articles/Pages/epistemic.html>)

⁵ Eliasmith, C., y Thagard, P. (2001), "Integrating structure and meaning: A distributed model of analogical mapping", *Cognitive Science* 25, 245-286.

⁶ Este trabajo fue realizado en el marco del proyecto de investigaci3n "El descubrimiento cient3fico desde las perspectivas de las reglas heur3sticas" dirigido por el Prof. Victor Rodriguez y subsidiado por FONCYT, Proyecto PICT 98, N3 04-04353.