

**teorema**

Vol. XVIII/2, 1999, pp. 33-48

## **Incompatibilidades en las teorías del contenido mental<sup>1</sup>**

Fernando Broncano

“Don’t let me be misunderstood”  
(interpretada por Nina Simone)

### ABSTRACT

We examine in this paper a family of current theories of mental content from the perspective of their commitment with three assumptions: the first is the representationalism, as it is exposed in the so-called Theory of Representational Mind; the second is the naturalism, as a constraint involved in many approaches to mind; and the third is the holism, as it is derived from the radical-interpretation theoretic approaches to mind and language. We argue that the most known theories are unable to meet jointly the three requirements. But we have strong reasons, on the other hand, to support each requirement one by one.

### RESUMEN

En este trabajo examinamos una familia de teorías actuales del contenido mental desde la perspectiva de su compromiso con tres supuestos: el primero es el representacionalismo, tal como lo encontramos en la llamada teoría de la mente representacional; el segundo es el naturalismo en cuanto restricción que se imponen algunos programas de investigación sobre la mente; el tercero es el holismo que proviene de las teorías de la mente y el lenguaje basadas en la interpretación radical como metodología. Nuestro examen muestra que las teorías más importantes han sido incapaces de cumplir los tres requisitos conjuntamente, a pesar de que tenemos razones muy fuertes para apoyarlos por separado.

Las teorías del contenido mental han surgido de la intersección de las teorías del significado y de las teorías de la mente, por tanto de una doble intención de explicar lo que el hablante sabe cuando usa o entiende el lenguaje y de la intención de explicar la acción voluntaria y compleja. Esta intersección convierte a estas teorías en un territorio de discusión metafísico que la semántica formal podía obviar hasta el momento refugiándose en el territorio de los objetos abstractos, pero al entrar en un contexto psicológico, cual es el de los contenidos, entramos sin poderlo evitar en el terreno de la ontología. En este territorio las teorías que refieren (o niegan la referencia) a entidades

explicativas están sometidas a constricciones que limitan las posibilidades de construcción teórica.

La tesis que presentamos es que existe un conflicto básico entre al menos tres objetivos de las teorías contemporáneas de la mente subyacente a buena parte de las discusiones contemporáneas y, lo que nos parece más interesante, que constriñe el campo de posibles teorías que uno puede desarrollar consistentemente. No podemos evitar la sospecha de que el conflicto sea irresoluble y básico. Que sea irresoluble porque se asiente sobre una imposibilidad de consistencia de estrategias programáticas, y que sea básico porque nos remita a estrategias asentadas en los convencimientos metafísicamente más profundos, aquellos que rozan el carácter de principios constitutivos. Constaré, sin embargo, que las teorías contemporáneas del contenido mental han sido incapaces de superar el conflicto, y que precisamente su existencia explica buena parte de las discrepancias entre las teorías.

Las tres ideas directrices constituyen los núcleos de programas de investigación con una larga tradición en la filosofía occidental y que han demostrado una gran capacidad de supervivencia. Llamaré a estos tres programas *el sueño de Descartes*, *el sueño de Hume* y *el sueño de Hegel*, pero mi exposición no será histórica, sino que subrayaré los rasgos contemporáneos que caracterizan el estado actual de estos programas.

#### I. EL SUEÑO DE DESCARTES: UNA CIENCIA REPRESENTACIONAL

Desde Descartes hasta Nelson Goodman ha persistido el proyecto de reconstruir completa y fundamentalmente el conjunto de nuestra experiencia mental del mundo. Esta reconstrucción ideal habría de realizar un ordenamiento de todos los contenidos mentales de forma que el resultado abarcara no solamente el contenido actual de nuestra conciencia sino el de todo contenido posible. En este proyecto constitucionalista no se establecen distinciones entre racionalistas ortodoxos cartesianos y fenomenalistas de origen lockeano o berkeleyano, ni tampoco, para nuestros propósitos, entre trascendentalistas y psicologistas. El hilo conductor es la idea de que la explicación de la acción humana exige la reconstrucción de la dinámica de los contenidos de la conciencia.

Desde mi punto de vista lo que define al cartesianismo es una doble característica: en primer lugar, el compromiso con la defensa de la intencionalidad como marca característica ontológica que divide el mundo de la física y el mundo de la mente, sea o no mente humana. Toda explicación contiene una premisa intencional que según este programa no es eliminable. La intencionalidad como marca de lo mental es, pues, una característica esencial del cartesianismo, pero no se agota en tal reconocimiento el cartesianismo. En se-

gundo lugar, la idea cartesiana entraña una explicación de la naturaleza de este carácter intencional que es la llamada *teoría representacional de la mente* (TRM) o teoría de las actitudes. El programa cartesiano considera coextensivo el carácter intencional de un suceso con su carácter representacional. El vehículo de la intencionalidad son las representaciones y las representaciones se individualizan por su contenido, en el que reside la intencionalidad. De manera que las teorías del contenido cartesianas presuponen la realidad de las representaciones, sea cual sea la naturaleza de éstas, entidades abstractas como las proposiciones o entidades mentales de carácter informacional o computacional. El doble compromiso cartesiano implica, por tanto, una taxonomía de actitudes suficiente al menos para explicar y predecir la conducta cotidiana de los sujetos y la tesis de que las representaciones son los vehículos de la intencionalidad.

El espectro de teorías que abarca el cartesianismo es amplio, de hecho agrupa a casi toda la tradición filosófica occidental y a la gran mayoría de las corrientes psicológicas vigentes, pues si algo ha demostrado el cartesianismo es su capacidad de persistencia en la historia, y ello se debe, desde mi punto de vista, a que tiene una base sólida de argumentos para subsistir. En primer lugar, porque apela a nuestro mundo cotidiano en el que la vida social demuestra que la teoría representacional de la mente está incorporada atrincheradamente en nuestras relaciones:

Las actitudes están entrelazadas en la estructura de las instituciones sociales, legales, políticas y otras. Nada podría ser un contrato o una invitación a cenar o una elección o una sentencia de muerte en ausencia de creencias, deseos e intenciones. Sin la atribución de actitudes proposicionales no habría justificaciones, excusas, ruegos o culpas a otros [Baker (1996), p. 4].

En segundo lugar, porque apela a los datos de la psicología contemporánea y las varias ciencias relacionadas. La TRM, para un numeroso grupo de psicólogos, formaría parte de nuestra dotación psicológica innata y la interpretación de la conducta ajena a través de las actitudes formaría parte de nuestras habilidades constitutivas. Este segundo argumento es empírico, pero sus consecuencias son fortísimas para la filosofía: una de ellas es que si se comprobase definitivamente que la TRM es constitutiva, en un sentido de necesidad nómica, *no podríamos pensar de otra manera* que a través de la psicología de actitudes.

En tercer lugar, como ha expuesto repetidamente Fodor, apela a la propia posibilidad de una psicología científica: a menos que se adopte la TRM resulta muy difícil que sea posible una ciencia psicológica nómica que pueda predecir las regularidades de la conducta humana, que tienen que incluir nuestras predicciones cotidianas, pues son las más robustas y más contrastadas en la práctica por la selección natural.

La persistencia del programa cartesiano podemos ahora entenderla por la fuerza de los argumentos en su favor y por el grado de exigencia de los criterios que debe cumplir cualquier teoría del contenido<sup>2</sup>. El cartesianismo ha sido posiblemente el blanco del mayor número de ataques en la filosofía contemporánea, pero tarde o pronto la fuerza de alguno de los principios que componen este cinturón protector ha hecho que el programa sobreviva a los ataques.

## II. EL SUEÑO DE HUME

El segundo de nuestros ideales es el de Hume, el de encontrar una ciencia de la naturaleza humana, pues Hume consideraba la búsqueda de una ciencia de la naturaleza humana no solamente como una extensión de la ciencia de la naturaleza, sino como parte del proyecto general científico. De hecho la actitud naturalista ante el contenido no es un mero lujo filosófico de materialistas militantes, es un proyecto que deriva de un lado de nuestra aspiración a una visión coherente del mundo en el que habitamos y de nuestro puesto en él, y de otro lado, como deseaba Hume, de la necesidad de conocer la naturaleza del hombre para fundamentar el valor del conocimiento.

El naturalismo no es un término unívoco, hay tantas formas de naturalismo como ciencias favoritas tiene cada uno, pero en la forma más estricta que adoptaremos aquí, se considera como una condición necesaria el que las propiedades explicativas que ocurren en las teorías de un dominio sean propiedades que puedan ser caracterizadas como propiedades naturales. El criterio de naturalidad es que una propiedad es natural si y sólo si es una propiedad que puede tener efectos causales. En la filosofía contemporánea se han establecido criterios más refinados para poder acoger en este criterio a las propiedades más usuales de las ciencias. En este criterio se incluye un supuesto de *cierre causal de las propiedades físicas*, en el sentido que Kim explica coloquialmente: “si uno toma cualquier evento físico y traza su trayectoria causal anterior o posterior, nunca se sale fuera del dominio físico” [Kim (1997a), p. 282]. La consecuencia es que, para no violar este principio, toda propiedad que tenga contenido empírico explicativo tiene que demostrarse que es directamente causal o que hereda la causalidad de los componentes en los que se realiza o, en último extremo, que sobreviene a una base física de realización.

Una vez que admitimos que este principio es correcto, las propiedades quedan divididas en propiedades que son directamente causales, propiedades que heredan las propiedades causales de su microestructura, y, en tercer lugar, propiedades de segundo orden [Kim (1997b)] que no son causales pero de las que puede establecerse claramente una base de sobrevenida. Muchas propiedades, como las funcionales, no son propiedades causales por sí

mismas sino que lo son en virtud de su base de sobreveniencia, pero de esta forma vicaria entran en la mayoría de las ciencias como propiedades explicativas con un grado suficiente de naturalismo.

El problema de la naturaleza del contenido se muestra así muy conectado con el problema más general de la naturaleza humana: lo que nos hace humanos es que nuestro comportamiento está causado por nuestras creencias y deseos y no por un orden frío de causas físicas. Y si lo que nos hace humanos es que nuestra acción está causada por el contenido de nuestra mente en tanto que agentes conscientes y libres, entonces una parte esencial de nuestra concepción humana y humanista consistirá precisamente en describir nuestra acción en términos causales. Desde esta perspectiva el proyecto de naturalizar el contenido es, pues, en cierto modo, independiente de que consideremos o no metodológicamente especiales las ciencias humanas, se trata más bien de la cuestión de dónde situamos los fenómenos de contenido en nuestro catálogo de propiedades naturales. De manera que, en tanto el programa de la naturalización del contenido no avance, nuestra aspiración a una visión unificada de la naturaleza estará insatisfecha.

### III. EL SUEÑO DE HEGEL

Por último, el holismo. El holismo es la gran aportación del idealismo a la cultura. Nace de la convicción metafísica de que lo que confiere identidad a los individuos concretos es el conjunto de relaciones que mantienen, y de que estas relaciones son el producto de un proceso histórico que no puede ser eliminado al dar cuenta de la naturaleza de tales individuos. El sueño del idealismo alemán fue encontrar una descripción de la naturaleza en la que las relaciones entre los objetos que componen el amueblamiento de la realidad fueran relaciones que establecieran una interdependencia profunda en la naturaleza de esos objetos. Pues lo que define y determina la identidad concreta de los objetos es, en la perspectiva idealista, la concreta determinación de sus relaciones, de manera que para conocer el mundo es preciso conocer el conjunto de las relaciones que se establecen en el tiempo y el espacio entre sus partes<sup>3</sup>.

La idea esencial del holismo es, pues, la de dependencia entre las partes de un sistema. Hay propiedades holistas en un sistema S cuando ocurre que el que un individuo posea o instancie una propiedad F que caracteriza al sistema sólo si otros componentes del sistema instancian ciertas propiedades (que pueden ser las mismas que F). Por ejemplo, decimos que la propiedad de pertenecer a una comunidad es holista porque el hecho de que una pertenezca a una comunidad implica que esa persona sigue reglas, y esta propiedad sólo es instanciable si otras muchas personas que pertenecen a la

comunidad la instancian también [Esfeld (1997)]. Uno de los problemas abiertos de la definición de holismo es el cómo caracterizar la relación de dependencia que se establece entre las partes de un sistema. Puede ser una dependencia metafísica, pero puede ser también solamente causal.

El problema del holismo es su difícil contención. Quienes comienzan diciendo que los términos sólo tienen significado en el contexto de la frase, pueden resbalar suavemente a la afirmación de que son las teorías las que confieren significado, para sostener seguidamente que en realidad son redes de teorías que se desarrollan en la historia, a lo que algunos pueden añadir que las teorías no son independientes de las prácticas y negocios con el mundo, y...<sup>4</sup>

Lo interesante de los tres programas anteriores es que persiguen ideas metodológicas que se constriñen mutuamente, pues una ciencia representacionalista no es ajena a la cuestión de naturalismo, un programa holista acerca de la cultura humana no es independiente de naturalismo, etc. Lo sorprendente es la constatación de que los programas contemporáneos en teorías del contenido mental se organizan alrededor de la suma de dos de las constricciones anteriores (el representacionalismo, el naturalismo, el holismo) pero excluyen positivamente la tercera. Es más, la exclusión de la tercera es precisamente lo que configura las señas de identidad del programa.

#### IV. FODOR O LA INCONSISTENCIA DEL HOLISMO CON LA CAUSALIDAD Y EL REPRESENTACIONALISMO

Las objeciones de Fodor representan la mayor fuente de dificultades para una teoría del contenido que pretenda ser holista al tiempo que representacionalista y naturalista. El naturalismo y el representacionalismo se exigen mutuamente: por una teoría naturalista entendemos una teoría que cumpla dos condiciones, la primera es que nos permita generalizaciones robustas, al menos de la misma fuerza que conseguimos en las generalizaciones de la psicología intencional cotidiana y, en segundo lugar, que sea compatible con la clausura causal de las propiedades físicas. Renunciar a estas dos condiciones sería renunciar a las dos características definatorias del naturalismo, una metodológica y otra ontológica.

Según Fodor, no todas las formas de holismo son igualmente dañinas. Desde 1983 Fodor viene sosteniendo que la contrastación o “fijación de la creencia” es holística, pues es uno de los ejercicios de la creatividad humana, de manera que se considera relativamente seguidor de Quine en lo que respecta a su filosofía de la ciencia, pero desde su punto de vista el problema está en el holismo de significado o de creencia. Si admitimos este holismo, afirma Fodor, debemos abandonar el programa naturalista en psicología.

El argumento más general contra el holismo de significado es que implica directamente la imposibilidad de una psicología predictiva. Si lo que determina el estado mental de un agente es el conjunto completo de su red de creencias, según Fodor se sigue directamente la imposibilidad de que dos agentes estén nunca en el mismo estado y de aquí la imposibilidad de generalizaciones nómicas o cuasi-nómicas en psicología puesto que no puede cuantificarse sobre estados mentales. El argumento afecta a muchas formas de holismo, pero es particularmente dañino para aquellas teorías del contenido entre cuyos objetivos esté el hacer compatible el representacionalismo y el naturalismo. Afecta, en particular, a la teoría del rol computacional, según la cual el contenido de un estado cognitivo está determinado por el conjunto sus relaciones computacionales con otros estados. El argumento aquí es que una psicología naturalista exige una identidad de contenidos robusta y la identidad robusta de contenido, afirma Fodor, no puede garantizarse si el contenido lo determinan todas las conexiones computacionales.

Hay dos defensas posibles de esta acusación de Fodor. La primera es introducir una función de similaridad de estados mentales como sustrato de las generalizaciones de la psicología intencional. Si los conceptos fuesen prototipos, y nuestro pensamiento fuera esencialmente a través de relaciones de similaridad, la identidad de estados podría garantizarse a través de una cuantificación sobre estados similares. La respuesta de Fodor a esta estrategia es doble: en primer lugar, afirma, existen contenidos sobre los que no puede establecerse ninguna función de esta clase. Fodor pone el ejemplo del contenido de “no ser un gato” [Fodor (1998), p. 101]: no hay nada que tengan en común una bañera y un teléfono, contenidos ambos que caen bajo el alcance de este estado, luego dos personas en las que el contenido de “no ser un gato” tenga un papel causal, no tienen sin embargo nada en común. En segundo lugar, su argumento más difundido es que una teoría de prototipos o cualquier otra noción basada en similaridades no podrá nunca ser composicional.

[...] la propuesta para componer prototipos se enfrenta al siguiente dilema: o bien trata las verdades lógicas (meramente) como casos extremos de verdades fiables estadísticamente, o admite que los pesos asignados a los rasgos en las matrices derivadas no son composicionales, *incluso si las matrices mismas lo son* [Fodor (1998), p. 103].

Esta última acusación es la que dirige contra la semántica de vectores propuesta por los Churchland: aunque los vectores en los espacios sean composicionales no lo son los rasgos primitivos con los que se configura el espacio.

La segunda estrategia de defensa del holismo deriva de la exigencia de un principio de caridad o principio de humanidad como condición de todas las generalizaciones de la psicología predictiva. Pero en este caso nos situa-

mos en las siguientes formas de incompatibilidad, pues esta estrategia implica el rechazo del programa naturalista porque conlleva directamente al anomalismo de lo mental propuesto por Davidson, o del programa representacionista, como ocurre con la versión del principio de caridad como principio metodológico de perspectiva, como sostiene Dennett.

#### V. LA INCONSISTENCIA DEL REPRESENTACIONALISMO CON EL NATURALISMO Y EL HOLISMO

La segunda forma de inconsistencia entre los tres programas es la que deriva de la unión del holismo con el naturalismo. El precio es el abandono del programa cartesiano de la TRM. La idea central es que quienes simultáneamente adoptan un compromiso holista y naturalista terminan abandonando la idea de un rico medio representacional con una estructura equivalente o isomórfica a la postulada por la psicología intencionalista de la TRM<sup>5</sup>. Este resultado se debe a que la TRM se vacía del papel central que tiene en todas las otras dos alternativas. Y este vaciamiento resulta de la conjunción de dos características aplicadas a los procesos cognitivos: la dependencia y la emergencia de propiedades causales.

Las formas más interesantes son las que se apoyan en los nuevos modos del computacionalismo, el conexionismo y las redes neuronales o en la neuro-computación, y las nuevas perspectivas de la llamada “vida artificial” (agentes autónomos o agentes situados o *animats*). Como ya es bien conocido, las redes neuronales procesan la información a través del mutuo ajuste de sus interconexiones o “pesos” de conexión. La capacidad de respuesta, y por consiguiente la capacidad de representación de las redes reside en la dinámica total del sistema que forma la red y posiblemente las conexiones de la red con otras redes. De esta forma desaparece la noción clásica de símbolo como soporte de la información y por consiguiente la relevancia del funcionalismo llamado de “máquina de Turing”. Son las singularidades que se producen en el espacio de fases de un sistema dinámico de una red las que adquieren relevancia semántica como portadoras de la información. El holismo de esta forma de ver el contenido nace inmediatamente del carácter físico de la red: los estados a los que atribuimos contenido son por su estructura estados relacionales de un sistema complejo que no puede ser analizado atómicamente.

Si lo que hace de los sistemas holistas sistemas holistas es la idea de *interdependencia*, lo que hace de los sistemas holistas sistemas naturalistas es la idea de *emergencia*. El holismo que tiene pretensiones naturalistas es el que propone que las propiedades de un nivel dado de organización de la naturaleza surgen de niveles de organización de la naturaleza que no poseen esas cualidades. Hay que notar que no todas las formas de emergentismo son



holistas, pero las formas de holismo que pretenden ser naturalistas deben serlo. Entiendo aquí por emergentismo la idea de que la organización de la naturaleza puede generar nuevas propiedades causales, en el sentido de propiedades de una clase diferente a la clase de propiedades que poseen los componentes del sistema y que, además, posee poderes causales. Desde las moléculas orgánicas autocatalíticas a las células, de las células a los organismos complejos y las especies, desde las neuronas simples a la capacidad de cálculos complejos, la autoorganización de la naturaleza habría producido la aparición de un montón de nuevas potencias causales. Pues el holismo naturalista está unido también a la idea de la autoorganización de la naturaleza<sup>6</sup>. A esta forma de holismo pertenecen una serie de términos que poseen ya un aura de novedad y cierto misterio: autocatálisis, autoorganización, caos, complejidad, sistemas no lineales, irreversibilidad, etc. Las tres familias más conocidas son el holismo conexionista y más generalmente neuroconexionista (Churchland, Clark) la teoría de los sistemas cognitivos como sistemas dinámicos (T. Van Gelder) y la teoría panselccionista (G. Edelman).

Hay una metáfora de Dennett [Dennett (1995)] que nos permite captar muy claramente esta mezcla de holismo y emergencia. Se trata de la idea de grúas en la evolución. Hay muchos ejemplos de tipos de grúas selectivas, pero aquella en la que Dennett está pensando prioritariamente es el lenguaje. El lenguaje como constructo social produce la mente y no a la inversa. Así, tendríamos sistemas evolutivos que se adaptan al medio, transformándolo y creando en él estructuras más complejas que, a su vez, realimentan la dinámica del sistema haciendo que éste evolucione hacia propiedades más complejas.

El sistema que tenemos que contemplar en esta construcción ya no puede ser un sistema interno, ni siquiera un sistema interno holístico, sino el sistema más complejo que forman las interacciones entre la estructura interna y la estructura del medio. No basta el simple contextualismo externalista, necesitamos una deixis esencial hacia tipos determinados de contextos que ya disponen de una estructura que realimenta la anterior. “Bailar un tango es cosa de dos”. No tienen ya sentido los procesos causales tomados aisladamente, porque nada nos indican sobre el orden y la estructura del proceso. Es el conjunto de la interacción el que da sentido a cada uno de los movimientos. En el holismo naturalista parece inevitable la suma del externalismo y el procesualismo [Clark y Chalmers (1998)]. El externalismo nos remite deícticamente a contextos particulares, y el procesualismo establece como unidad de cuenta para todo propósito cognitivo, procesos temporales que implican un orden asimétrico temporal.

El compromiso con la emergencia de nuevas propiedades causales tiene una explicación muy clara: para escapar a la amenaza del argumento de la pobreza de estímulo que acabó con el asociacionismo clásico piagetiano, es necesario el recurso a las grúas dennettianas: en la interacción dentro del sis-

tema, hay una parte que necesariamente debe ser más compleja, sea cual sea la medida de complejidad que usemos y que asimétricamente es la fuente de la complejidad de la otra parte. Y esta interacción debe ser causal, de ahí el componente necesariamente emergentista del holismo naturalista. Es la riqueza del medio la que introduce complejidad física en el proceso de arborización neuronal que, a su vez, permite la riqueza representacional suficiente para asimilar nueva información [Sejnowski y Quartz (1997), Clark y Karmiloff-Smith (1993), Clark y Thorton (1997)].

En la agenda del holismo naturalista no está por supuesto el programa representacionalista cartesiano. Y no puede estarlo por una razón que me parece patentemente clara: la TRM, para expresarlo en términos hegelianos, nunca podrá ser más que un momento en la dialéctica del proceso. No es necesario que el eliminativismo sea el destino necesario, como es el caso flagrante de los Churchland, las representaciones siempre serán el término con el que designamos la complejidad interna de una parte del sistema heterogéneo, pero el lenguaje en el que hablamos del proceso no será representacional, aún si es compatible con la psicología natural.

Dejaremos a un lado la difícil cuestión de si esta forma de emergencia se compromete con una forma de causación hacia abajo [Kim (1997a y 1997b)]. La respuesta depende estrechamente de cómo se describa el sistema emergente: si logramos que la emergencia sea una emergencia mereológica las nuevas propiedades causales se heredan de las propiedades causales de las partes sin que la novedad causal introduzca ningún elemento sobreviniente. Pero eso nos compromete a abandonar la sobrevenida de las propiedades mentales. En otro caso el holismo naturalista se encuentra con graves problemas similares a los del funcionalismo del rol causal.

## VI. LA INCONSISTENCIA DEL NATURALISMO CON EL HOLISMO REPRESENTACIONALISTA

Un numeroso grupo de filósofos, para terminar, prefiere mantenerse fiel a una forma de representacionalismo y al mismo tiempo sostener formas fuertes de holismo. Como representacionalistas, aunque sea *sui generis*, sostienen que explicar la conducta exige adscribir al sujeto actitudes proposicionales, como holistas sostienen que estas actitudes reciben el contenido en virtud de su inserción en una red de actitudes. Esta cooperación conduce a fuertes enfrentamientos con el programa naturalista, es decir con la idea de que esta explicación es al tiempo una explicación causal. La tensión entre estas dos premisas con la tercera es uno de los más trágicos fracasos de la filosofía contemporánea. El éxito del programa hubiera significado establecer un puente entre las ciencias humanas y las ciencias naturales. El producto de un

argumento explicativo debería individualizar una y sólo una cadena causal. Pero esta explicación sufre de varios problemas, mortales cada uno de ellos: la explicación en términos de razones es múltiplemente instanciable en diversas cadenas causales, la causa mental es pre-vaciada por la causa física, y, en el mejor de los casos, implica causación hacia abajo. No es extraño que algunos autores primigeniamente funcionalistas, sea en la forma de una teoría del contenido como rol causal o formas computacionalistas más abstractas, hayan derivado hacia formas de antinaturalismo abiertas. El caso paradigmático es el de Hilary Putnam, aunque no es difícil encontrar otros.

Esta forma de inconsistencia entre dos programas depende estrechamente de la validez de una cierta forma kantiana de argumentación trascendental. La naturaleza trascendental de la argumentación holista se pone de manifiesto cuando reparamos que al explicar una acción el observador se enfrenta a ilimitadas posibles explicaciones que no puede eliminar en virtud solamente de cuestiones de hecho, ni siquiera en virtud de cuestiones de significado, porque precisamente son las que debe decidir. Éste es el problema que detecta el llamado contexto de traducción radical. El argumento trascendental que efectúa el teórico consiste en presentarnos un objeto que tiene una propiedad de hecho, pero que solamente puede tenerla de derecho en virtud de su pertenencia a un marco más general que tiene otra cierta propiedad. Es una forma de presuposición trascendental que presupone una forma de holismo de marco. Tomemos como ejemplo el caso del holismo de interpretación de Davidson en "Belief and the Basis of the Meaning": si nos tomamos en serio los contextos de interpretación radical como marcos desde los que hay que establecer una teoría del significado, nos vamos a encontrar con que las creencias, los significados y la conducta (verbal, asertiva) van los tres juntos en un mismo paquete a la hora de adscribir, de determinar cada uno de ellos. El holismo que nos presenta Davidson es el que nos encontramos al resolver una ecuación con tres incógnitas: necesitamos conocer el valor de dos de las variables si queremos resolver el de la tercera, pero, para desgracia nuestra, la naturaleza no nos proporciona un sistema de ecuaciones independientes, de manera que necesitamos algún principio externo para determinar el valor de las incógnitas. Es precisamente lo que nos suministra el Principio de Caridad: nos tomamos a nosotros mismos como referentes y entonces fijamos los valores de las creencias, o valores de verdad, en términos similares a grandes rasgos a como haríamos nosotros mismos. El Principio de Caridad se convierte de esta manera en una condición *a priori* sin la cual no podemos adscribir propiedades como el significado a las palabras o sentido a las acciones.

El holismo de Putnam pertenece a esta clase de holismo de estructura trascendental. Su campo de argumentación es, sin embargo, mayor que el de Davidson y, por supuesto, Quine. Se trata de un holismo en el que las consideraciones ontológicas, semánticas y epistemológicas van juntas, en el que

las argumentaciones semánticas presuponen que ya hemos fijados los valores de nuestras creencias y que hay una aportación del mundo. El argumento básico desarrollado a partir de *Truth, Reason and History* es que la causalidad presupone la relevancia y la relevancia presupone una elección condicionada por las creencias, es decir, la ontología permanece muda sin la epistemología. De manera que el holismo de Putnam está basado en una estructura de argumentación trascendental pero su alcance implica la mutua dependencia de ontología, epistemología y semántica. Al igual que en el caso del Principio de Caridad de Davidson o Quine, en Putnam existe una referencia esencial, *a priori*, a nuestro esquema conceptual sin la cual seríamos incapaces de determinar o bien los significados, o bien las creencias, o bien las relaciones causales y la aportación del mundo.

Una respuesta posible a nuestra presentación es que estos programas no son más cartesianos que naturalistas, si bien es cierto que son antinaturalistas también lo es que abandonan la idea cartesiana de lo mental. Pero ésta es una conclusión precipitada debido a que confundimos el cartesianismo y el atomismo. No es necesario para ser cartesianos postular representaciones dotadas de contenido como entidades separadas, aunque muchos cartesianos de hecho lo hagan; por el contrario, podemos postular que la representacionalidad de los estados mentales deriva de propiedades holísticas del sistema mental, o si se quiere, está localizado en un nivel personal. Es solamente cuando pretendemos que nuestra explicación es causal cuando las representaciones dejan de ser explicativas. Lo que es esencial es que consideremos que las creencias y deseos cumplan una función robusta explicativa y que esta explicación es una explicación última y completa, que no necesitamos descender a otros niveles para rellenarla con algo que faltaría en el nivel interpretativo.

Lo que un intérprete completamente informado puede aprender acerca del significado del hablante es todo lo que hay que aprender; y lo mismo sirve para lo que hablante cree [Davidson (1983), p. 315, citado en Child (1994) p. 1].

Las bases de argumentación son distintas según los autores; el argumento del anomalismo de lo mental de Davidson es tal vez el más conocido: si una acción exige razones, nada que no sea una creencia puede contar como razón. Otros argumentos pueden ser más sofisticados, como son aquellos que apelan a las condiciones de inteligibilidad o interpretabilidad de las acciones como [McDowell (1994)]. La estructura trascendental deriva del hecho de que un suceso físico como una acción o una preferencia deben ser situados en el “espacio de las razones”, para usar el término sellarsiano que McDowell ha vuelto a poner de moda, y esta situación induce un abandono definitivo de la idea de que una explicación naturalística añadiría algo a la explicación de la acción.

## VII. DESCONTANDO PROPIEDADES DE UNA TEORÍA DEL CONTENIDO

La situación en la que nos encontramos recuerda en cierto modo las antinomias kantianas. Las antinomias surgen cuando tenemos razones para sostener algo pero también las tenemos para sostener lo contrario, a lo que habría que sumar el que, además, tenemos enfrente propiedades que tienen un carácter último. Sospecho que esta es la situación y que los filósofos deberíamos acostumbrarnos a la idea de la incompatibilidad entre los tres ideales, de modo que vamos a tener que optar entre propiedades descontando aquéllas que por unas u otras preferencias nos parezcan más prescindibles. No he podido demostrar que la incompatibilidad vaya más allá que el hecho de que las teorías actuales son incompatibles, pero no es difícil diagnosticar la fuente de las dificultades. El naturalismo sostiene como un núcleo esencial el principio del *cierre causal de la realidad*: si viajamos hacia delante o hacia atrás en la cadena de las causas de un evento nunca salimos del espacio físico. El cartesianismo, del mismo modo, sostiene que hay un principio de *cierre en el espacio de las razones*: nada que no posea las propiedades representacionales puede aspirar a contar como fundamento o consecuencia racional de una representación. El holismo, también, tiene su principio de cierre, el *cierre de dependencias para todo proceso real*: nada que no posea las dependencias adecuadas puede aspirar a entrar en un proceso dinámico de la realidad. La realidad es dinámica porque las cosas son interdependientes. Cuando unimos dos de estos tres principios de cierre, el tercero se nos convierte en un territorio completamente ajeno al mundo del contenido que queremos reconstruir.

Pero debemos recordar que el precio de un dualismo en nuestra visión del mundo es extremadamente alto, no porque metodológica o epistémicamente sea insoportable, sino porque convierte nuestro puesto metafísico en el orden de las cosas en una pregunta sin respuesta. Lo mismo cabe decir del abandono del representacionalismo, induce una reforma de tal calibre en el conjunto de nuestro esquema conceptual que cabe sospechar que el otro posible esquema simplemente sea incomprensible para nosotros. Parecería que entonces el eslabón más débil es el holismo, que es relativamente poco costoso prescindir de él, pero es una conclusión errónea: convertimos inmediatamente en máximamente misteriosas las propiedades dinámicas de lo mental. Y todavía tenemos que explicar cómo es posible el aprendizaje sin suponer, como Platón en el *Menón* que ya lo sabemos todo, y el aprendizaje es difícilmente explicable sin una concepción relacional de los estados cognitivos.

*Departamento de Filosofía, Lógica y Filosofía de la Ciencia  
Universidad de Salamanca  
Campus Unamuno, Edificio F.E.S., E-37007, Salamanca*

E-mail: [broncانو@gugu.usal.es](mailto:broncانو@gugu.usal.es)

NOTAS

<sup>1</sup> La redacción de este trabajo se ha beneficiado del proyecto de investigación PB95-0125-C06-04 de la DGICYT. Versiones previas fueron leídas en 1996 en el simposio del CSIC sobre H. Putnam y en el Segundo Congreso de la Sociedad Española de Filosofía Analítica. Quiero agradecer los comentarios a estas tesis de Hilary Putnam, Manuel Liz, Diana Pérez, Sandra Lazzer, Lizza Skydelsky, Marcelo Sabatés y Jesús Vega.

<sup>2</sup> El cartesianismo no está comprometido con una teoría única del contenido pero sí con una lista de condiciones que toda teoría debería cumplir. Aunque los grados de exigencia son desigualmente rigurosos, me parece que estas condiciones serían: (i) *discriminación del contenido*, que se mide por su sensibilidad a la formulación lingüística, (ii) *independencia de las condiciones causales*, (iii) *existencia de negación*, (iv) *el criterio de generalidad de Evans*: un sistema irrepresentacional debe cumplir el criterio de que si asigna a los  $x$  la propiedad  $Fx$  y los  $x$  resultan ser de un mismo tipo que los  $y$  entonces es capaz de asignar  $Fy$  [Evans (1982), p. 75], (v) *la primera persona como criterio epistémico de conocimiento de la representación* (vi) *capacidad de productividad y sistematicidad*. Los cartesianos insisten de manera determinante en uno de estos, considerando que demarcan conceptualmente la propia naturaleza de TRM (por ejemplo, el criterio de composicionalidad), pero no es inconcebible pensar en que deben cumplirse en bloque, ni tampoco es inconcebible el que coexistan con posibles compromisos de realización mutua, sin que ello implique el abandono de una TRM más o menos ortodoxa

<sup>3</sup> El holismo y la construcción intensional de las relaciones están profundamente relacionados. No es por casualidad que el atomismo lógico emplease la construcción extensional de las relaciones como su arma más afilada contra el idealismo presente en la cultura europea. Y no es por casualidad tampoco que muchos críticos del atomismo lógico crean que la construcción extensional no es más que una forma verbal de resolver un problema metafísico.

<sup>4</sup> Lo interesante del holismo es que se puede presentar bajo formas y grados de fuerza diferentes. El holismo ontológico de Leibniz, por ejemplo, no tiene excesivas consecuencias metodológicas, mientras que en el extremo contrario el holismo de Wittgenstein es esencialmente metodológico sin adoptar compromisos ontológicos. Las formas de holismo se pueden ordenar por la pretensión de su alcance: los holismos metodológicos se limitan a la confirmación o contrastación de creencias, otros afectan al significado de las creencias son los holismos semánticos y cognitivos: otros, más allá, insertan la semántica en un mundo no conceptual de prácticas corporales, hasta llegar a quienes no separan las propiedades representacionales de *affordances* o propiedades ecológicamente prefabricadas para ser representadas. Cabe, incluso, un holismo de tipo hegeliano de comunidad entre todo lo existente, es lo que Dennett ha llamado el *holismo Zen* [Dennett (1998), p. 219].

<sup>5</sup> Las teorías que caen bajo esta forma de rechazo no son todas las teorías holistas, quisiera aclarar antes de exponer mi argumento. Algunas teorías del rol funcional, sea rol computacional o rol causal en términos abstractos, son positivamente holistas y no abjuran de un compromiso naturalista. Pero hoy sabemos que tienen problemas por ambos

lados del compromiso. Desde el punto de vista de su compromiso holístico, no hay respuestas aceptadas como válidas a la objeción de que no son capaces de generar desde dentro una propuesta de similaridad de dos mentes diferentes lo suficientemente robusta para sostener las generalizaciones de la psicología intencional y, desde el punto de vista naturalista, tienen el problema de que el contenido concebido en esos términos es impotente causalmente. Consideraremos, pues, que son teorías que se encuentran actualmente ante un dilema que seguramente les obligará a decantarse hacia opciones de mayor compromiso con una u otra de las alternativas que estamos presentando aquí. Lo mismo podemos afirmar respecto a las teorías teleológicas del contenido, sean de carácter evolutivo o de teleología de diseño. Son teorías que no tienen tan claro su componente holístico. De hecho no son incompatibles con una teoría fuertemente representacionista, aunque pueden no serlo. Jacob ha propuesto recientemente que las teorías teleológicas deberían adscribirse abiertamente al holismo [Jacob (1997), cap. 6], en cuyo caso también evolucionarían hacia el tipo de incompatibilismo que estamos tratando.

<sup>6</sup> Algunos autores han señalado ya que estas formas de holismo emergentista se encuentran ya en los estratos más básicos de organización de la materia, en particular en la mecánica cuántica: la superposición de los sistemas cuánticos ha sido entendida por algunos autores como una muestra de propiedades que emergen de propiedades básicas pero que una vez generadas ya no podemos rastrear las propiedades componentes microfísicas [Humphreys (1996)]. En este trabajo se alude a algunos conocidos resultados en mecánica cuántica como los experimentos de *Aspect*. Otros ejemplos que se han puesto de sistemas holísticos es el de los sistemas con dinámicas dirigidas por atractores extraños, en los que la propiedad emergente es epistémicamente irreducible a las partes [Newman (1996)].

#### REFERENCIAS BIBLIOGRÁFICAS

- BAKER, L. R. (1996), *Explaining Attitudes. A Practical Approach to the Mind*, Cambridge, Cambridge University Press.
- CARRUTHERS, P. (1996), *Theories of The Theory of Mind*, Cambridge, Cambridge University Press
- CLARK, A. (1998), "Where Brain, Body, and World Collide", *Daedalus*, vol. 127, pp. 257-80.
- CLARK, A. y KARMILOFF-SMITH, A. (1993), "The Cognizer's Innards: A Psychological and Philosophical Perspective of the Development of Thought", *Mind and Language*, vol. 8, pp. 487-519
- CLARK, A. y THORNTON, C. (1997), "Trading Spaces: Computation, Representation and the Limits of Uniformed Learning", *Behavioral and Brain Sciences*, vol. 20, pp. 57-90.
- CLARK, A. y CHALMERS, D. (1998), "The Extended Mind", *Analysis*, vol. 58, pp. 7-19.
- CHILD, W. (1994), *Causality, Interpretation and the Mind*, Oxford, Clarendon Press.
- CHURCHLAND, P. (1993), "Fodor and Lepore: State Space Semantics and Meaning Holism", *Philosophy and Phenomenological Research*, vol. 53, pp. 679-82.

- (1998), “Conceptual Similarity Across Sensory and Neural Diversity: The Fodor/Lepore Challenge Answered”, *The Journal of Philosophy*, vol. 95, pp. 5-32.
- DAVIDSON, D. (1983), “A Coherence Theory of Truth and Knowledge”, en Henrich, D. (ed.) (1983), *Kant oder Hegel?*, Stuttgart, Klett-Cotta; reimpresso en Lepore, E. (1986), *Truth and Interpretation. Perspectives on the Philosophy of Donald Davidson*, Oxford, Basil Blackwell, pp 307-19.
- DENNETT, D. (1995), *Darwin’s Dangerous Idea*, Nueva York, Bantham.
- (1998) *Brainchildren*, Cambridge Mass., MIT Press.
- ESFELD, M. (1998), “Holism and Analytic Philosophy”, *Mind*, vol. 107, pp. 365-79.
- HUMPHREYS, P. (1996), “Aspects of Emergence”, *Philosophical Topics*, vol. 24, pp. 53-70.
- JACOB, P. (1998), *What Minds Can Do*, Cambridge, Cambridge University Press.
- KIM, J. (1997a), “The Mind-Body Problem: Taking Stock after Forty Years”, en Tomberlin, E. (ed.) (1997), *Philosophical Perspectives*, vol. 11, *Mind, Causation and World*, Atascadero, CA, Ridgeview.
- (1997b), “Does The Problem of Mental Causation Generalize”, *Proceedings of the Aristotelian Society*, vol. 98, pp. 281-97.
- MCDOWELL, J. (1994), *Mind and World*, Cambridge, Mass., Harvard University Press.
- NEWMAN, D.W. (1996), “Emergence and Strange Attractors”, *Philosophy of Science*, vol. 63, pp. 245-61.
- QUARTZ, S. y SEJNOWSKY, T. (1997), “The Neural Basis of Cognitive Development: A Constructivist Manifesto”, *Behavioral and Brain Sciences*, vol. 20, pp. 537-96.