

REALISMO INTENCIONAL Y NATURALIZACIÓN DEL CONTENIDO*

GUIDO VALLEJOS

Universidad de Chile

El presente trabajo tiene como propósitos centrales examinar y evaluar tres teorías filosóficas que, desde sus distintas perspectivas, pretenden fundamentar el realismo intencional. Se entiende por realismo intencional aquella posición filosófica que plantea que la intencionalidad intrínseca es parte del orden natural. Las tres teorías realistas actuales en las que se centra este trabajo intentan dar una visión naturalista del contenido intencional de las representaciones mentales. Dos de las teorías examinadas —sostenidas por Fodor y Millikan— exhiben una fuerte tendencia reduccionista. El presente artículo intenta mostrar que dicha tendencia no aporta mucho a la comprensión del rol de las estructuras semánticas en los organismos que las poseen. Consecuentemente, se considera una tercera visión naturalista del contenido —sostenida por Dretske— que, sin caer en excesos reduccionistas, abre una senda más promisoría para una semántica naturalizada.

INTRODUCCIÓN

La mayoría de las disciplinas que forman parte del marco teórico que configura lo que se ha dado en llamar ciencia cognitiva conciben la mente como un mecanismo de procesamiento de la información. De acuerdo a esta concepción, el procesamiento es llevado a cabo usando símbolos como base. Las operaciones de procesamiento son caracterizadas como computacionales, esto es, como transformaciones ejecutadas sobre los símbolos que portan la información. Los símbolos en cuestión son caracterizados como representaciones. Dichas representaciones poseen una sintaxis, que es crucial para la ejecución de los procesos computacionales, y una semántica, en virtud de la cual esas representaciones tienen un contenido.

Como se afirma que los símbolos tienen rasgos representacionales, su contenido debe ser acerca de algo externo a los símbolos mismos. Aun cuando el problema del contenido representacional y de su intencionalidad es un tema controversial en la filosofía de la ciencia cognitiva (véase, por ejemplo, Cummins 1989), la concepción general de la mente que, en la opinión de la mayoría de los filósofos, los científicos cognitivos sostienen todavía parece plausible, al menos para el propósito que cumplen estas consideraciones introductorias. De acuerdo a esa concepción general (cf., por

*El autor desea agradecer al profesor Peter Carruthers del Departamento de Filosofía de la Universidad de Essex por sus valiosos comentarios y sugerencias al presente trabajo.

ejemplo, Stillings et al. 1987), la mente es un mecanismo procesador de la información cuyos estados son caracterizados intencionalmente. Sin embargo, las teorías cognitivas influidas por el modelo computacional de la mente están mayormente interesadas en encontrar las leyes o algoritmos que expresan los patrones que gobiernan las computaciones ya sea en máquinas o mentes. Consecuentemente, las teorías computacionales se interesan también en describir la arquitectura cognitiva en la cual las leyes o algoritmos así formulados pueden guiar las computaciones. El énfasis en los procesos computacionales y en la arquitectura cognitiva nos lleva a concluir que, aun cuando gran parte de los científicos cognitivos supone que los signos internos o representaciones tienen rasgos representacionales y, en consecuencia, un contenido intencional, ellos no proporcionan una explicación al problema de cómo el contenido de una representación mental puede ser acerca de, y determinado por, su relación con algo en el mundo. Quizás el problema de la intencionalidad del contenido representacional, tal como lo acabamos de formular, implica una discusión metafísica en la cual el científico cognitivo no pretende involucrarse.

En este sentido, P.N. Johnson-Laird, un teórico partidario de la visión computacional de la mente, caracteriza la ciencia cognitiva como sigue:

La ciencia cognitiva —haciéndome eco de las palabras que William James pronunciara hace cerca de un siglo— es el estudio de las mentes individuales finitas, y supone que hay un mundo físico en el espacio y en el tiempo, y que los procesos mentales capacitan a las personas para percibir ese mundo, para tener pensamientos y sentimientos y para estar conscientes de sí mismos. Estas suposiciones pueden ser puestas en duda, pero, como lo señala James, su discusión se llama metafísica y cae fuera del alcance de la ciencia... La ciencia procede haciendo suposiciones simplificadoras acerca del mundo y explicando el cuadro idealizado resultante... (1988a: 387).

De acuerdo a la cita, la explicación de cómo el contenido de una representación puede ser acerca de, y determinado por, su conexión con algo en el mundo conlleva una discusión en torno a la mayoría de las que Johnson-Laird llama 'suposiciones simplificadoras' y, podríamos agregar, no explicadas, sobre las cuales gran parte de las teorías en ciencia cognitiva construyen su visión computacional de la mente. Una consideración del problema de la intencionalidad del contenido de las representaciones implicaría, por ejemplo, una discusión acerca de las características de los elementos del mundo a los cuales las representaciones mentales están conectadas y también definir el tipo de relación que las representaciones tienen con aquellos elementos. Si Johnson-Laird está en lo correcto, puede concluirse que el problema de la intencionalidad del contenido representacional es un problema de carácter filosófico o, al menos, un problema que algunos científicos cognitivos consideran que es materia de preocupación exclusiva por parte de los filósofos. Si vamos a considerar dicho problema desde el punto de vista de la filosofía, podemos formularlo de la siguiente manera: Dando por supuesto el hecho de que hay signos internos que son representaciones y que esas representaciones tienen la propiedad de ser acerca de sus representados, el problema es cómo formular las condiciones bajo las cuales se lleva a cabo la relación de intencionalidad entre dichas representaciones y los elementos del mundo a los cuales ellas representan. Tal formulación debería poner de manifiesto, además, cómo el contenido de una representación se determina si dichas condiciones son satisfechas.

Encontrar una solución al problema planteado no ha sido una tarea fácil para los filósofos. En el pasado se han hecho intentos para formular condiciones del tipo descrito más arriba, pero dichos intentos se han visto frustrados por el hecho de que la formulación de las condiciones en cuestión parece requerir una apelación a nociones intencionales. Aparte del problema obvio de circularidad que tal situación provoca, este modo de abordar el problema del contenido intencional ha llevado a algunos filósofos a creer que las nociones semánticas y, en general, las nociones intencionales, forman un círculo cerrado. Como lo señala Fodor 1990, la creencia de que las nociones semánticas e intencionales forman un círculo cerrado puede llevar a la creencia de que dichas nociones son irreducibles a nociones naturalistas y esto, a su vez, puede llevar a la creencia de que los ámbitos de lo intencional y de lo natural forman dos conjuntos disjuntos de propiedades. La consecuencia de todo esto es que las disciplinas que operan con nociones intencionales no se conciben como formando parte del corpus de las ciencias naturales. Algunos filósofos contemporáneos ligados a los desarrollos de la ciencia cognitiva han pensado que el círculo en la caracterización de las nociones intencionales puede romperse mediante la naturalización de la intencionalidad. El quiebre del círculo en cuestión tendría como resultado el terminar con el dualismo entre lo natural y lo intencional.

Hay dos resultados extremos producto de estos intentos de naturalizar la intencionalidad:

- a) o bien se concluye que no hay en el orden natural algo como la intencionalidad intrínseca, aun cuando puede aceptarse, teniendo a la vista la utilidad que obtenemos al predecir el comportamiento de algunos organismos, que podemos asumir pragmáticamente la perspectiva intencional y hacer adscripciones intencionales a esos sistemas;
- b) o bien puede concluirse que hay algo como la intencionalidad intrínseca —en tanto opuesta a una intencionalidad derivada de propósitos prácticos, como en el caso anterior— que es parte del orden natural.

Podemos llamar a la primera de estas posiciones Irrealismo Intencional y a su opuesta Realismo Intencional. En este trabajo nos centraremos en algunas variedades del realismo intencional en conexión con el problema del contenido representacional. Entenderemos por realismo intencional aquella posición que sostiene que el orden natural exhibe algo como la intencionalidad intrínseca. Si este planteamiento básico del realismo intencional se conecta con el problema de la intencionalidad del contenido, según lo hemos caracterizado más arriba, tendríamos como resultado que el realismo intencional haría posible la formulación de las condiciones bajo las cuales el contenido de una representación es determinado en términos naturalistas.

El propósito de este trabajo es examinar y evaluar tres teorías que tratan de entregar una visión naturalista del contenido fundadas en el realismo intencional. Con la finalidad de considerar críticamente algunas de las variedades del realismo intencional, asumiremos que dicha posición general es al menos una postura teórica plausible y defendible. En consecuencia, no argüiremos contra aquellos intentos de naturalización que tienen como resultado la desaparición de la intencionalidad intrínseca del orden natural.

En primer lugar, examinaremos la teoría causal y fiscalista de Fodor (1987, 1990).

La solución que propone Fodor al problema de la naturalización del contenido representacional consiste en reducir la relación intencional, mediante la cual el contenido es determinado, a una ontología de propiedades fisicalistas y de relaciones causales entre tales propiedades. La aproximación causal de Fodor está inspirada por las teorías semánticas informacionales (entre otras, por la propuesta por Dretske 1981). Este tipo de semántica tiene el problema de ser incapaz de explicar los casos de error o de falla en la representación. La solución propuesta por Fodor toma como punto de partida este tipo de semántica, pero la reformula de tal manera de permitir la explicación de los casos de error. Al examinar el intento de naturalización de Fodor, trataremos de mostrar que su proyecto no permite tener una visión del contenido como un fenómeno natural. Nuestra conclusión será que la teoría de Fodor está dirigida más bien a fundar una psicología intencional en una ontología fisicalista que, en su opinión, es común a todas las ciencias naturales. Como veremos, para Fodor la noción de contenido es crucial para una psicología intencional que tenga la aspiración de ser compatible con los rasgos semánticos que la psicología cotidiana adscribe a las actitudes proposicionales. Ahora bien, él necesita mostrar que la psicología intencional puede ser una ciencia natural. Fodor piensa que aquella parte de la psicología intencional que se preocupa de los procesos y estados mentales puede ser articulada en términos computacionales. La concepción computacional de la mente, debido a su compromiso con la neutralidad ontológica del funcionalismo, es compatible con la ontología fisicalista de las ciencias naturales. El gran problema surge para la psicología intencional cuando debe considerar los rasgos semánticos de las actitudes proposicionales. Como la noción de contenido es una noción intencional que tradicionalmente se ha mostrado renuente a ser reducida a términos naturalistas, la psicología intencional aparece como una ciencia que está fuera de la esfera de las ciencias naturales y, en consecuencia, como proporcionando una base de sustentación para el dualismo ontológico. Por lo tanto, parece ser necesario mostrar que esta noción problemática puede reducirse a la ontología básica de las ciencias naturales. Sin embargo, lo anterior no permite concebir el contenido como un fenómeno natural; es decir, no permite visualizar el rol causal que le cabe al contenido en los organismos cuyo comportamiento está de alguna manera determinado por el uso que hacen de las estructuras semánticas. La reducción de Fodor solamente muestra que la psicología intencional es compatible con la ontología de las ciencias porque una de sus nociones problemáticas —i.e. la de contenido— es reducible a esa ontología.

En segundo lugar, examinaremos el intento de Millikan de entregar una visión naturalista del contenido desde una perspectiva teleológico-evolucionista. Trataremos de poner de manifiesto que, para los efectos de articular una teoría naturalista de la intencionalidad, para Millikan es suficiente mostrar que los elementos intencionales se ajustan al marco categorial de, y su caracterización está fundada en, las hipótesis básicas de una ciencia como la biología. Los elementos intencionales pueden ser caracterizados teleológicamente, esto es, pueden expresarse en términos de aquello 'para lo cual sirven' o, lo que viene a ser lo mismo, en términos de la función que un elemento intencional se supone debe ejecutar. Al hacer tal caracterización, el teórico de orientación teleológica puede, como Millikan lo hace, asumir un punto de vista evolucionista basado en la historia selectiva de los elementos intencionales en consideración. Sin embargo, las teorías teleológicas no tienen necesariamente que apelar a la historia selectiva para caracterizar la función de un elemento. Por ejemplo, Fodor pensó en

algún momento (manuscrito editado por Lycan 1990) que había algo como una teleología naturalizada en la cual las apelaciones a las rutas históricas del diseño evolucionista no eran necesarias y donde las funciones podían caracterizarse usando contrafácticos (e.g. la función de un elemento 'e' es aquella para la cual dicho elemento hubiese sido seleccionado aun cuando tal y tal hubiese sido el caso). Es conveniente señalar en este punto que lo que aquí hemos llamado irrealismo intencional puede también asumir un punto de vista teleológico y evolucionista (véase, por ejemplo, Dennett 1987). De acuerdo a esta visión irrealista, no hay una diferencia cualitativa entre nosotros cuando diseñamos un artefacto como un voltímetro o un medidor de gasolina y le asignamos la función de indicar un cierto estado en el mundo, y la Madre Naturaleza cuando, por decirlo de alguna manera, asume la perspectiva intencional y diseña un organismo como una función indicadora de similar carácter que ella considera beneficiosa para la supervivencia de tal organismo.

Después de examinar la concepción de Millikan acerca de las representaciones, concluiremos que si su teoría es considerada como un intento de naturalizar el contenido intencional —o de presentar una visión naturalista de la intencionalidad, para usar sus propios términos—, dicho intento se muestra como insuficiente porque cae en circularidad. Su caracterización de la relación entre el contenido representacional y el correspondiente objeto intencional, especialmente en el caso de los organismos que están en los peldaños más altos de la escala evolutiva, exhibe demasiadas nociones intencionales, lo que hace a su teoría sospechosa de circularidad.

Teniendo presente que los dos intentos de naturalización que examinaremos fracasan en algunos aspectos, a nuestro juicio, cruciales, consideraremos una tercera opción. Este tercer intento de naturalización es el sostenido por Dretske 1988. En nuestra opinión, la propuesta de Dretske tiene varias virtudes. En primer lugar, concilia una semántica informacional —o semántica de los indicadores, como la llama Block 1986— con lo que llamaremos una semántica de los mecanismos interpretativos. En segundo lugar, puede decirse que su intento de naturalización no es reductivo, ni en el sentido ontológico de Fodor ni tampoco en el sentido biológico de Millikan. Su teoría puede caracterizarse como un esfuerzo de hacer compatible la explicación del rol causal del contenido con algunas hipótesis básicas y algunos datos proporcionados por la biología evolucionista y la etología. A pesar de que la teoría del contenido propuesta por Dretske merece una consideración crítica más detallada que la que se hará en este trabajo, será suficiente para nuestros propósitos el poder mostrar en forma más o menos clara que la visión de Dretske abre una opción, quizás no nueva, pero al menos promisoria, para la consideración del problema que nos ocupa. El proyecto de naturalización del contenido que encontraremos en Dretske no es reductivo ni eliminativista con respecto a lo intencional (como es el caso de Quine y de Churchland); tampoco es reductivo combinado con un antirrealismo respecto de lo intencional (como es el caso de Dennett); no es un proyecto reductivista que evita la desaparición de la intencionalidad del orden natural (como en la reducción ontológica de Fodor); finalmente, no intenta reducir los elementos intencionales al marco categorial de la biología (como en el sentido más débil de reducción que encontraremos en Millikan). Quizás, y esa será la moraleja de este trabajo, podremos tener una concepción más clara del contenido como un fenómeno natural si somos capaces de apelar a algunas hipótesis de determinadas ciencias naturales que son en algún sentido compatibles con algunas de nuestras

intuiciones filosóficas acerca del significado, tratando, al mismo tiempo, de evitar cualquier tentación reductivista.

1. LA TEORÍA DEL CONTENIDO DE FODOR

1.1. *Teoría representacional de la mente y semántica realista*

No es fácil establecer en qué lugar de la teoría general de la mente sostenida por Fodor surge el problema de la naturalización del contenido. Por esta razón, es conveniente comenzar esta primera sección con una visión general de su teoría con el objeto de establecer en forma más o menos clara dónde surge el problema en cuestión y cuáles son los rasgos que lo definen. Distinguiremos nueve fases que, a nuestro juicio, caracterizan en sus rasgos esenciales la concepción de la mente de Fodor. Al final de esa caracterización esperamos haber circunscrito en forma clara el lugar que le cabe al problema de la naturalización del contenido en la teoría de Fodor.

En términos generales, los aspectos significativos de la teoría de la mente de Fodor pueden caracterizarse como sigue:

1) El punto de partida es la suposición de que el marco general al cual las aproximaciones teóricas a la mente deben adaptarse es la psicología cotidiana. El hombre común explica el comportamiento apelando a actitudes proposicionales. Por ejemplo, Juan abrió el refrigerador porque deseaba beber agua mineral y porque creyó que había agua mineral en el refrigerador.

2) El patrón explicativo usado por la psicología cotidiana muestra que hay una relación causal entre estados mentales y entre éstos y la reducción del comportamiento. Por otra parte, las actitudes proposicionales atribuidas a los sujetos en este tipo de explicaciones aparecen exhibiendo un contenido que es evaluable semánticamente. Por ejemplo, respecto de este último punto, el deseo de Juan de beber agua mineral puede ser satisfecho si él puede encontrar una botella del líquido en el refrigerador, y su creencia de que hay agua mineral en el refrigerador es verdadera solamente si, de hecho, él puede encontrar una botella del líquido al abrir el refrigerador; si así no fuera, su deseo no sería satisfecho y su creencia sería falsa.

3) La psicología cognitiva aparece como sustentando aquella parte de la psicología cotidiana que dice relación con las conexiones causales entre estados mentales, y entre éstos y el comportamiento. Este tipo de psicología concibe la mente como un mecanismo procesador de información, donde los estados mentales son fases identificables en, y producidos por, secuencias de procesos mentales.

4) Una concepción de la mente como un mecanismo de procesamiento requiere de un medio en el cual dicho procesamiento se efectúe. El medio en cuestión es proporcionado por la hipótesis de que hay un conjunto de representaciones mentales —un lenguaje del pensamiento o *Mentalés*— en el cual los procesos mentales se llevan a cabo (Fodor 1975).

5) Dada la hipótesis del lenguaje del pensamiento (LP), una actitud proposicional puede caracterizarse como una relación funcional que un organismo tiene con una representación mental, y los procesos mentales pueden concebirse como instanciaciones (*tokenings*) de representaciones mentales (Fodor 1985, 1987). En esta fase de la teoría de Fodor, las representaciones mentales son caracterizadas como símbolos que

expresan proposiciones y cuyos rasgos más relevantes son los de carácter sintáctico. Fodor concibe los rasgos sintácticos como rasgos físicos de segundo orden que definen la forma (*shape*) de un símbolo. De acuerdo a esto, Fodor afirma que los procesos mentales son computaciones, esto es, transformaciones en virtud de la sintaxis de los símbolos de Mentalés. Las transformaciones sintácticas alteran la forma de los símbolos, i.e. sus rasgos físicos abstractos; si esto es así, los procesos computacionales pueden pensarse como secuencias causales de instanciaciones de representaciones mentales. Como se dijo, las representaciones mentales son símbolos que expresan proposiciones. Las transformaciones sintácticas que ocurren en el procesamiento mental reflejan las relaciones semánticas de entrañe entre las proposiciones expresadas por los símbolos involucrados en el procesamiento.

6) Lo que se dijo en 5) expresa en términos muy generales la Teoría Representacional de la Mente (TRM) de Fodor. Esta teoría concibe la mente como una máquina impulsada sintácticamente. En TRM la sintaxis desempeña el rol de mediar entre las propiedades causales de los símbolos y las propiedades semánticas de las proposiciones expresadas por esos símbolos. En este contexto, las propiedades semánticas que se consideran, como ya se dijo, son las relaciones de entrañe entre las proposiciones expresadas por los símbolos de Mentalés.

7) TRM puede hacerse más específica mediante la formulación de una taxonomía funcional de los procesos mentales, la cual puede considerarse como la expresión de la arquitectura cognitiva de la mente. Fodor presenta esta taxonomía funcional en *The Modularity of Mind* (1983). Allí, él examina los niveles de la arquitectura cognitiva, cómo están organizados, y cómo cada uno de esos niveles opera en cada una de las fases del procesamiento de la información.

8) La hipótesis del lenguaje del pensamiento, TRM y la tesis de la modularidad de la mente (TMM) son compatibles con, y sustentan gran parte de, la concepción psicológica cotidiana que dice relación con las conexiones causales entre estados mentales. De acuerdo a las teorías más arriba mencionadas, el rol causal de los estados mentales se establece en concordancia con su rol funcional en una secuencia de procesamiento. Pero LP, TRM y TMM son partes de una psicología computacional y, como tales, aun cuando asumen que la mente es un sistema intencional, son incapaces de explicar de qué manera los estados mentales tienen un contenido intencional. En otras palabras, partiendo del supuesto de que la mente es un sistema intencional, estas teorías de la psicología computacional pueden solamente explicar de qué manera este sistema opera en virtud de la forma de sus elementos simbólicos y en virtud de las características de su arquitectura. Sin embargo, son incapaces de explicar cómo este mecanismo formal —o esta máquina impulsada sintácticamente, para emplear la expresión de Fodor— es alimentada por contenidos semánticos. En suma, las teorías en cuestión no pueden explicar cómo las representaciones mentales tienen un contenido. Como partes de una psicología computacional LP, TRM y TMM están limitadas por lo que Fodor 1981 llama solipsismo metodológico. Según esta posición, la única opción metodológica para una psicología computacional es abstraer de las representaciones todos aquellos rasgos que conectan dichas representaciones con propiedades en el mundo y concentrarse exclusivamente en la forma de las representaciones.

9) Lo que se necesita, entonces, es una teoría del contenido que se ajuste al marco impuesto por la psicología cotidiana. Tal teoría debe asumir el punto de vista de una

psicología intencional y explicar cómo los estados mentales, según los concibe la psicología computacional, pueden tener un contenido. Si así no fuera, uno de los rasgos principales que la psicología cotidiana adscribe a las actitudes proposicionales quedaría sin explicación, a saber, su evaluabilidad semántica. Contrariamente a lo que algunos psicólogos y psicolingüistas afirman (cf., por ejemplo, Johnson-Laird 1983, 1988a y 1988b, y Jackendoff 1983 y 1987), la psicología computacional no puede proporcionar una teoría adecuada del contenido porque es incapaz de explicar cómo las representaciones mentales están conectadas a ítems en el mundo. Siendo esto así, la psicología computacional no puede dar cuenta de la intencionalidad o de la propiedad que poseen las representaciones de ser acerca de algo. Además, una psicología que intentara entregar una visión del contenido desde un punto de vista computacional caería inevitablemente en lo que aquí hemos llamado irrealismo intencional, puesto que no reconocería la relación entre una representación y aquello que ésta representa como una relación real; es decir, como aquella relación en virtud de la cual el contenido de una representación es efectivamente determinado.

De acuerdo a lo que se expresó en los nueve puntos anteriores, es claro que el problema de la intencionalidad del contenido no puede resolverse en el seno de una psicología computacional. Es importante señalar que el problema de la naturalización no surge al interior de una concepción computacional de la mente. En efecto, al asumir la neutralidad ontológica del funcionalismo, la concepción de los procesos y estados mentales que este tipo de psicología propone tiene una realizabilidad universal, esto es, las computaciones pueden realizarse en cualquier tipo de materia. Siendo esto así, la visión computacional de la mente es compatible con la ontología de las ciencias naturales. Lo anterior se manifiesta en forma más clara cuando se considera el rol que desempeñan en los procesos mentales los rasgos sintácticos de los símbolos que participan en dichos procesos, a saber, hacer posible que las cadenas de transformaciones sintácticas puedan concebirse como secuencias causales (véase 5), 6) y 7) arriba).

El problema del contenido no parece ser crucial en una concepción como TRM, claramente influida por un patrón explicativo funcionalista. Una teoría como TRM es útil para explicar las diferencias entre creencias, deseos e intenciones, pero no sirve para diferenciar, por ejemplo, la creencia de que P es el caso de la creencia de que Q es el caso. Fodor 1987 expresa esta limitación de una teoría semántica orientada por el funcionalismo de la siguiente manera:

Supongamos que el 'silogismo práctico' es más o menos verdadero (como, después de todo, seguramente lo es). Lo que éste dice es que si queremos que P y si creemos que no-P a menos que Q, entonces, siendo todo lo demás igual, tratamos de hacer que Q sea el caso, para cualquier P y Q. Así formulada, dicha generalización pretende aplicarse a todas las creencias y deseos, sin importar cuáles sean sus contenidos. De este modo, la taxonomía de estados mentales que dicha generalización implica, distingue entre el creer y el desear, pero no entre el creer que tal y tal del creer esto y aquello. (1987:70)

Suponiendo que Fodor está en lo correcto al afirmar que la perspectiva funcionalista opera para explicar el procesamiento mental como un factor productivo de los estados mentales, pero no sirve para establecer las condiciones bajo las cuales el contenido es determinado, la pregunta que surge es por qué el problema del contenido intencional es tan crucial para una psicología intencional. La única respuesta plausible que hemos

sido capaces de encontrar en Fodor es que la importancia del problema radica en que la psicología intencional tiene la tarea de explicar la evaluabilidad semántica de las actitudes proposicionales. Si una psicología intencional es capaz de dar cuenta de aquel rasgo de las actitudes proposicionales, entonces será compatible con, y proporcionará una base de sustentación teórica para, la concepción psicológica cotidiana.

La pregunta que surge en este punto es por qué el contenido tiene que ser naturalizado. La respuesta a esta pregunta está contenida en su mayor parte en lo que, respecto de este punto, ya expresáramos en la Introducción de este trabajo. Una formulación de las condiciones de determinación del contenido de una representación que use nociones intencionales llevaría a revivir la vieja creencia filosófica de que el ámbito de lo intencional forma un círculo de nociones que no puede expresarse en términos de otras nociones que no sean las intencionales. Como se dijo, la consecuencia de lo anteriormente enunciado es la reafirmación del viejo dualismo según el cual hay dos conjuntos disjuntos de propiedades en el mundo: las naturales y las intencionales. En la opinión de Fodor, el dualismo ontológico lleva al irrealismo respecto de lo intencional. Si esto es así, entonces la noción de contenido necesita ser objeto de una reducción naturalista. Los resultados de esta estrategia reductiva serán en su mayor parte beneficiosos para la psicología intencional porque harán posible que este tipo de psicología aparezca como una ciencia natural.

Sin embargo, es necesario señalar que las preocupaciones de Fodor en este caso son, en su mayor parte, de carácter ontológico, y que él no está mayormente interesado en mostrar de qué manera el contenido puede concebirse como un fenómeno natural. El problema, tal como lo plantea Fodor, puede caracterizarse en forma más clara diciendo que es un problema de ontología de las ciencias. En efecto, hay una noción que es central para la psicología intencional, en este caso la de contenido, y esta noción es, además, problemática, puesto que se ha mostrado renuente a ser reducida a otras nociones que no sean intencionales. Esto serviría como base para afirmar que la psicología intencional está fundada en un tipo de propiedades distinto al tipo sobre el cual están fundadas las ciencias naturales. En la opinión de Fodor, todas las ciencias naturales pueden reducirse a una ontología de propiedades fisicalistas abstractas y de relaciones nomológico-causales entre esas propiedades. Sin embargo, esto no significa que cada una de las ciencias especiales desaparezca en tal reducción. Por ejemplo, la geología no se convierte en física en virtud de tal reducción, pero aparece como fundada en este substrato ontológico básico. Lo mismo sucedería con la psicología intencional si la noción de contenido pudiera ser reducida a esta ontología de propiedades fisicalistas y relaciones nomológico-causales.

De acuerdo a lo expuesto, la reducción del contenido a relaciones nomológico-causales entre propiedades fisicalistas abstractas tiene como consecuencia el fundar la psicología en un substrato ontológico que es común a todas las ciencias naturales. De esta manera, este tipo de psicología no aparece como sustentando un dualismo de propiedades ni tampoco apelando a un tipo especial de causalidad (como, por ejemplo, la causalidad intencional propuesta por Searle 1983). Así, la intencionalidad aparece ocupando un lugar en el orden natural y, en consecuencia, la ciencia que opera con nociones intencionales es una ciencia natural. Si es posible mostrar que la intencionalidad intrínseca tiene un lugar en el orden natural, entonces la tesis del realismo intencional con respecto al contenido es, al menos, una posición teórica plausible. En lo

que sigue, presentaremos la naturalización que del contenido propone Fodor y que, según lo dicho, podría contar como un argumento en favor de la tesis del realismo intencional.

1.2. *Una teoría causal del contenido*

El problema de la intencionalidad surge en la teoría de Fodor cuando es necesario especificar las condiciones bajo las cuales el contenido de un símbolo no-lógico y primitivo de Mentalés (como 'caballo', para usar uno de los ejemplos favoritos de Fodor) es determinado. Tales condiciones deben especificarse en términos no-semánticos y no-intencionales. Fodor piensa que al dar tal especificación es posible sentar las bases para una semántica naturalista.

Para Fodor, la mejor solución al problema de la naturalización del contenido es una semántica de carácter causal e informacional. De acuerdo a este tipo de teorías, el contenido de una representación expresa aquellos factores causales que hacen posible la producción de instanciaciones de esa representación. Siendo esto así, el contenido de un símbolo porta la información acerca de los contenidos causales que han producido sus instanciaciones. La relación informacional entre el contenido de un símbolo y los contextos causales que produjeron las instanciaciones de dicho símbolo está fundada nomológicamente.

Si conectamos la tesis general de una teoría causal, tal como la esbozamos más arriba, con el modo en que Fodor formula el problema del contenido, tendríamos que decir que el símbolo primitivo y no-lógico de Mentalés 'caballo' expresa la propiedad *caballo* porque hay una relación de covarianza nómica entre instancias de la propiedad *caballo* e instanciaciones del símbolo 'caballo'. Una solución causal o covariacionista al problema del contenido no requiere, en opinión de Fodor, una especificación de los mecanismos que sustentan la relación de covarianza nómica entre instancias de propiedades en el mundo e instanciaciones de símbolos de Mentalés. Si tal relación puede sostenerse en forma confiable, entonces es semánticamente irrelevante el establecer qué tipos de mecanismos son los que median entre los dos extremos de dicha relación. Por lo tanto, lo que es relevante para una semántica naturalizada es que se pueda establecer una relación de covarianza confiable —esto es, nomológica y contrafácticamente sustentada— entre instancias de propiedades e instanciaciones de símbolos, sin importar cómo esta relación esté mediada. Es conveniente detenerse en esta última observación ya que algunos autores, en particular Cummins 1989, ven las teorías covariacionistas cayendo en circularidad. Los autores en cuestión visualizan este problema porque piensan que, al tratar el problema de la determinación del contenido, dichas teorías hacen una apelación implícita a mecanismos inferenciales que utilizan conocimiento representacional almacenado y, en consecuencia, una apelación a la noción de representación, que es, precisamente, la que el teórico covariacionista pretende explicar. Es muy difícil encontrar esta apelación tácita, explícitamente negada por Fodor, en la teoría del contenido que estamos examinando. Es más, Fodor pone como una de las condiciones de adecuación de su teoría el que sea aplicable a la posibilidad conceptual —y no nomológica— de una mente que posea un solo símbolo susceptible de ser instanciado. Al poner esta condición de adecuación para su teoría, Fodor está tratando de obtener una teoría atomista del contenido. Una semántica

atomista no toma en consideración, como una condición para la determinación el contenido de una representación, las relaciones actuales o posibles que el contenido de dicha representación tenga con los contenidos de otras representaciones. Estas consideraciones muestran que cualquier referencia al contenido de otras representaciones, como una condición que determina el contenido, se evita en forma explícita.

A pesar de que una teoría causal parece ser, para Fodor, la mejor solución al problema de la naturalización del contenido, este tipo de teoría debe solucionar un problema que es crucial. La dificultad mencionada se pone de manifiesto cuando se examina una formulación usual de la teoría causal:

... las instanciaciones de un símbolo denotan sus causas, y el símbolo tipo expresa la propiedad cuyos especímenes causan en forma confiable las instanciaciones de dicho símbolo. (1987:99)

Que los especímenes de una propiedad causen en forma confiable las instanciaciones de un símbolo significa que hay una relación nomológica que sustenta la conexión causal entre un espécimen de la propiedad, digamos, *caballo*, y la propiedad de ser una instanciación del símbolo 'caballo'. Según esto, una condición suficiente para que el símbolo 'caballo' exprese la propiedad *caballo* es que es nomológicamente necesario que solamente especímenes de la propiedad *caballo* causen instanciaciones de 'caballo'. Pero es evidente que especímenes de propiedades que no son *caballo* pueden causar instanciaciones del símbolo 'caballo'. Una vaca en la oscuridad o en un día nublado puede causar una instanciación de 'caballo'. Un teórico causalista podría decir, al considerar este problema, que, puesto que solamente especímenes de *caballo* causan de manera confiable instanciaciones de 'caballo', una instanciación de dicho símbolo causada por vacas es evidentemente un error. Sin embargo, de acuerdo a la versión usual de la teoría causal, una condición suficiente para que un símbolo exprese una cierta propiedad es que sea nomológicamente necesario que solamente instancias de esa propiedad causen instanciaciones de ese símbolo. Siendo esto así, las instanciaciones de 'caballo' causadas por vacas no expresan la propiedad *caballo*, sino más bien la propiedad disyuntiva *caballo o vaca*. En otras palabras, dado que o bien instancias de *caballo* o bien instancias de *vaca* son suficientes para causar instanciaciones del símbolo 'caballo', entonces, de acuerdo a la formulación corriente de la teoría causal, el símbolo 'caballo' expresa la propiedad de ser un caballo o de ser una vaca. Según esto, la versión usual de la teoría causal tiene el problema de ser incapaz de explicar los casos de error o de representación fallida. Al asignar contenido disyuntivo a símbolos cuyas condiciones suficientes, en lo que a causalidad se refiere, son disjuntas, no puede explicar los casos de falsas instanciaciones de un símbolo. Lo que se ha expuesto corresponde a lo que Fodor llama el problema de la disyunción.

Para Fodor 1990, el problema de la disyunción surge al interior de las teorías causales porque dichas concepciones son incapaces de diferenciar la información que las instanciaciones de un símbolo portan del significado de dicho símbolo. Fodor expresa este punto como sigue:

Los errores hacen surgir el problema de la disyunción, pero el problema de la disyunción no es, en el fondo, un problema acerca del error. El problema de la disyunción, en el fondo, consiste realmente en la diferencia entre significado e información... La información está ligada a la etiología de un modo en que el significado no lo está... Por contraste, el significado

de un símbolo es una de las cosas que todas sus instanciaciones tienen en común, cualquiera sea el modo en que éstas sean causadas... De esta manera, la información sigue a la etiología y el significado no, y ésa es la razón por la que tenemos el problema de la disyunción si identificamos el significado de un símbolo con la información que portan sus instanciaciones. (1990:90)

Para Fodor, entonces, al considerar el problema de la disyunción es necesario explicar cómo aquello que él llama el 'carácter robusto' del significado puede obtenerse a pesar de la heterogeneidad de relaciones informacionales en que pueden entrar las instanciaciones de un símbolo. En otras palabras, una teoría causal del contenido debe explicar cómo el significado de un símbolo puede permanecer sin alteraciones a pesar de la variedad de tipos de relaciones causales actuales y posibles que pueden producir instanciaciones de ese símbolo. Más concretamente, instanciaciones de 'caballo' pueden ser causadas de muchas y diferentes maneras, pero a pesar de esta diversidad de causas, el símbolo 'caballo' significa o expresa la propiedad *caballo*. Según esto, no solamente instancias de la propiedad *caballo* causan instanciaciones del símbolo 'caballo', pero es suficiente que algunas instancias de *caballo* causen de manera confiable instanciaciones de 'caballo' para poder obtener el carácter robusto que ostenta el significado. Dicha característica se obtiene al seleccionar una relación causal semánticamente relevante, y mostrando que el resto de las relaciones causales heterogéneas es asimétricamente dependiente de la primera relación causal seleccionada.

Si el carácter robusto del significado puede obtenerse de la manera anteriormente señalada, y volviendo al ejemplo de representación fallida, entonces resultará que las instanciaciones de 'caballo' causadas por vacas son asimétricamente dependientes de instanciaciones de 'caballo' causadas por instancias de la propiedad *caballo*. Esto significa que las instanciaciones de 'caballo' tienen que llevar información acerca de caballos, de otra forma, como lo señala Fodor, "no llevarían información acerca de nada" (1990:91). En otras palabras, si la conexión vaca-'caballo' es rota, esto no afecta la conexión caballo-'caballo'; pero si la conexión caballo-'caballo' es rota, entonces la conexión vaca-'caballo' tendría que romperse también.

Hemos introducido la propuesta de Fodor en términos bastante vagos y generales en este punto y necesitamos, con el objeto de hacer justicia al planteamiento de este autor, complementar la presentación de su teoría con un breve examen de los supuestos ontológicos que respaldan el planteamiento de la dependencia causal asimétrica como una solución al problema de la disyunción. Fodor afirma que el mobiliario ontológico básico del mundo son propiedades fisicalistas y abstractas. El problema de la naturalización del contenido intencional se reduce, entonces, al problema de articular una teoría que establezca las condiciones suficientes bajo las cuales un conjunto de propiedades es acerca de otro conjunto de propiedades. Más concretamente, el establecer bajo qué condiciones, por ejemplo, el símbolo 'caballo', perteneciente a Mentalés, puede ser acerca de, expresar o significar la propiedad *caballo*. La posición de Fodor es que es suficiente, para que el símbolo 'caballo' exprese la propiedad *caballo*, que algunos especímenes de dicha propiedad causen instanciaciones de 'caballo' y que el resto de las conexiones causales entre instanciaciones de 'caballo' e instancias de propiedades diferentes a *caballo* sean asimétricamente dependientes de instanciaciones de 'caballo' causadas por caballos. De esta manera, la relación causal semánticamente relevante es la

existente entre especímenes de *caballo* e instanciaciones del símbolo 'caballo'. La pregunta que surge en este punto es de qué manera fundamentar una generalización del tipo: " 'caballo' significa *caballo* porque es suficiente que instancias de *caballo* causen instanciaciones de 'caballo', e instanciaciones de 'caballo' que son causadas por instancias de propiedades que no son *caballo* son asimétricamente dependientes de instanciaciones de 'caballo' causadas por caballos", dado el hecho de que no solamente caballos causan instanciaciones de 'caballo'. La respuesta de Fodor será que esta generalización está sustentada por contrafácticos. Fodor afirma que:

... las generalizaciones sustentadas por contrafácticos son (o bien idénticas a, o bien) respaldadas por leyes causales, y las leyes son relaciones entre propiedades. (1990:93)

De acuerdo al contenido de la cita anterior, podemos reformular la generalización enunciada más arriba como sigue: " 'caballo' significa *caballo* porque hay una relación nomológica entre la propiedad de ser un caballo y la propiedad de ser una causa de instanciaciones de 'caballo', y las relaciones nomológicas entre propiedades que no son *caballo* y la propiedad de ser una causa de instanciaciones de 'caballo' son asimétricamente dependientes de la primera relación nomológica".

Fodor piensa que es un hecho básico indesmentible que el orden natural contiene propiedades y relaciones nomológicas entre esas propiedades. Afirma que no es una buena idea analizar las relaciones nomológicas en términos de verdades contrafácticas acerca de individuos. Si lo anterior es una mala idea a juicio de Fodor, la pregunta que surge en forma inmediata es la siguiente: ¿Cómo sabemos, entonces, que estamos seleccionando la relación nomológico-causal semánticamente relevante, respecto de la cual las otras relaciones nomológicas involucradas en las instanciaciones de un determinado símbolo son asimétricamente dependientes? En opinión de Fodor, para responder esta pregunta no necesitamos saber demasiado acerca de los contrafácticos que hacen verdaderas estas relaciones nomológicas. Dada la relación nomológica que sustenta la conexión entre especímenes de *caballo* e instanciaciones de 'caballo', y dado que hay relaciones nomológicas entre propiedades distintas de *caballo* e instanciaciones de 'caballo' que son asimétricamente dependientes de la primera, todo lo que necesitamos saber es que si la relación caballo-'caballo' fuera rota, entonces las conexiones entre no-caballos e instanciaciones de 'caballo' también deberían desaparecer. Fodor piensa que para destruir su teoría sería necesario encontrar un mundo posible cercano (o nomológicamente posible) en el cual la relación nomológica entre, digamos, la propiedad *caballo* y la propiedad de ser una causa de instanciaciones del símbolo 'caballo' no pueda sostenerse. Fodor expresa este punto como sigue:

... solucionar el problema [del contenido] requiere proporcionar condiciones suficientes para la intencionalidad, y no condiciones suficientes y necesarias. De modo que si se quiere argüir contra las conclusiones metafísicas de este trabajo, se debe construir un mundo en el cual mis contrafácticos estén todos en su lugar pero donde X no significa lo que yo digo que significa. (1990:96)

Si todos los supuestos ontológicos que Fodor considera indesmentibles fueran fáciles de asimilar y si alguien aceptara las condiciones que él pone para la refutación de su teoría, entonces dicha teoría aparecería como imbatible. No argüiremos contra los detalles y supuestos ontológicos de la teoría de Fodor; tampoco intentaremos presentar

contraejemplos, que ya suman docenas, procurando seguir las instrucciones de este filósofo. Como se dijo en la Introducción, estamos más interesados en evaluar la plausibilidad de algunos intentos de mostrar de qué manera el contenido puede concebirse como un fenómeno natural que en los detalles de cada una de las propuestas que estamos considerando.

Con el objeto de tener una idea de cómo Fodor enfrenta las objeciones que se formulan a su teoría creemos que será suficiente considerar cómo enfrenta una objeción simple. Hasta donde sabemos, no ha habido instanciaciones verdaderas del símbolo 'unicornio'. Sin embargo, el hecho de que no haya habido instanciaciones verdaderas de un determinado símbolo no significa que es falso que la propiedad *unicornio* esté ligada nomológicamente con la propiedad de ser una causa de instanciaciones de 'unicornio'. Fodor piensa que la dependencia causal asimétrica puede expresarse para este caso como sigue: si no hubiese unicornios en algún mundo posible cercano al nuestro (esto es, un mundo nomológicamente posible; es decir, un mundo donde las relaciones nomológicas entre propiedades son similares a nuestro mundo actual o mundo de referencia), entonces las instanciaciones del símbolo 'unicornio' causadas por instancias de propiedades que no sean unicornios no serían posibles en nuestro mundo. Lo anterior no significa otra cosa sino que ha habido falsas instanciaciones de 'unicornio' en nuestro mundo, porque, si hubiera habido unicornios en él, entonces éstos habrían causado instanciaciones de 'unicornio'. El único fundamento para esta afirmación es que podemos encontrar mundos posibles cercanos —es decir, nomológicamente similares al nuestro— en los cuales podemos postular la existencia de unicornios. La dependencia causal asimétrica, de acuerdo a lo dicho, se sostiene en virtud de la apelación a contrafácticos y a mundos posibles en los cuales los contrafácticos aplicables a nuestro mundo son también aplicables.

El examen que hemos hecho de la teoría de Fodor exhibe claramente, en nuestra opinión, su carácter ontológico. La noción de contenido es reducida a una ontología fisicalista. A pesar de que mediante esta reducción no es mucha la claridad que obtenemos acerca del contenido como un fenómeno natural, al menos hay que conceder que lo que se ha obtenido es afincar una noción fundamental y problemática de la psicología intencional en la ontología de las ciencias naturales. Consecuentemente, la reducción cumple el propósito de mostrar que este tipo de psicología forma parte de las ciencias naturales. Sin embargo, la impresión final es que, en lugar de naturalizar la noción de contenido, Fodor está tratando de solucionar un problema atingente a la ontología de las ciencias. En la próxima sección de este trabajo consideraremos otro tipo de aproximación al problema de la naturalización del contenido intencional propuesto por R.G. Millikan (1984, 1986, 1989). Contrariamente a la aproximación ontológica de Fodor, Millikan trata de caracterizar los elementos intencionales, incluido el contenido, en términos de una ciencia especial: la biología.

2. MILLIKAN Y LA CATEGORIZACIÓN BIOLÓGICA DEL CONTENIDO REPRESENTACIONAL

2.1. *Una categorización biológica de los elementos intencionales*

La propuesta de Millikan puede ser caracterizada como un intento de mostrar que el lenguaje y la mente son categorías biológicas. Sin embargo, esta forma de tratar el problema de la naturalización del ámbito de lo intencional no emerge de consideracio-

nes científicas. Lo que Millikan hace es tomar como punto de partida ciertas hipótesis muy generales acerca del comportamiento de los mecanismos biológicos, las cuales se ajustan al patrón explicativo evolucionista. Posteriormente, ella trata de mostrar que los elementos intencionales se ajustan al patrón de comportamiento que exhiben los mecanismos biológicos, según se describen en las hipótesis básicas que sirven de punto de partida.

En lo que sigue entregaremos una visión general de algunos de los elementos que articulan la teoría de esta autora. Es conveniente señalar que la teoría de Millikan exhibe tal multitud de detalles, a veces confusos, que cualquier intento de dar una visión general de su teoría puede ser considerado como incompleto.

Millikan afirma que es plausible decir de un mecanismo biológico que tiene una función propia en virtud de la cual ha sido seleccionado en el proceso de evolución. Por ejemplo, el corazón tiene como función propia el bombear sangre al organismo, y esto da a los organismos que poseen dicho mecanismo una ventaja selectiva. Aparte de bombear sangre, hay otras cosas que el corazón hace y que pueden concebirse como funciones de ese mecanismo. Por ejemplo, el corazón late a un ritmo regular. No obstante, no tiene sentido decir que en el curso de la evolución dicho mecanismo fue seleccionado para ejecutar la función de latir en forma regular. Lo que realmente confiere un valor de supervivencia a los sistemas que poseen tal mecanismo y lo que ha hecho posible la proliferación del mismo en aquellos sistemas que lo poseen, ha sido precisamente lo que se ha caracterizado como su función propia.

Aunque podemos identificar un mecanismo biológico por su función propia, ello no implica que dicho mecanismo siempre, o la mayoría de las veces, desempeñe esa función propia. El mecanismo puede estar afectado por una malformación, u otras partes del sistema al cual pertenece el mecanismo pueden exhibir defectos, o, finalmente, las condiciones ambientales que rodean al organismo no son las más propicias para la ejecución adecuada de la función propia de un mecanismo. De lo anterior puede concluirse que un mecanismo biológico puede ejecutar su función propia solamente cuando las condiciones para tal ejecución son normales. La noción de normalidad no es, en este caso, una de carácter estadístico, sino más bien normativo. Para distinguirla de la noción estadística, Millikan usa en mayúsculas el término 'Normal'.

Las condiciones Normales para la ejecución de la función propia de un mecanismo biológico son parte de lo que Millikan llama la explicación Normal de por qué ese mecanismo tiene determinada función como su función propia. Una explicación Normal registra los factores históricos que han posibilitado la selección y proliferación de esa función. Dicho tipo de explicación muestra por qué tal función ha sido seleccionada como ventajosa para la supervivencia de un determinado sistema biológico en el curso del proceso de evolución. El término 'Normal', que se aplica tanto a las condiciones para la ejecución de una función propia como a las explicaciones que muestran por qué tal función ha sido seleccionada como ventajosa para un determinado sistema no es, como se dijo, una noción estadística, sino más bien una noción que se establece teniendo presente los factores históricos reales —y no posibles— en los cuales la función se ejecutó con un alto valor adaptativo para un sistema biológico dado. Una función puede haber sido ejecutada muy pocas veces en el pasado y tener, no obstante, un alto valor adaptativo para los individuos de una especie que poseen tal mecanismo. Por ejemplo, los individuos masculinos de la especie humana pueden producir millones de esperma-

tozoides, pero hay muy pocos de ellos que pueden ejecutar la función propia de fecundar el óvulo, posibilitando así la propagación de la especie. Es, entonces, el alto valor adaptativo que ciertos mecanismos han tenido en la ejecución de su función propia, y no la frecuencia en la ejecución de esa función, lo que ha hecho posible la proliferación de esos mecanismos al interior de determinada especie.

Con el objeto de conectar este tipo de explicación histórica con las leyes naturales, Millikan afirma que una explicación Normal para la ejecución de una función propia debe concordar con la operación de dichas leyes. En otras palabras, una explicación Normal no solamente necesita tener un carácter histórico, sino que, además, debe ser compatible con las leyes naturales y con las instancias de tales leyes.

Millikan piensa que el marco categorial de funciones propias, condiciones Normales y explicaciones Normales, fundado en la hipótesis de que los mecanismos biológicos son seleccionados y proliferan por el valor de supervivencia que tiene la ejecución de su función propia para los organismos que los poseen, puede aplicarse al análisis del lenguaje y de la mente. Si tal aplicación es posible, entonces los mecanismos lingüísticos e intencionales pueden considerarse como subsumidos bajo categorías biológicas. Siendo esto así, entonces es posible articular una teoría naturalista del lenguaje y de la mente. En lo que sigue presentaremos en forma breve la teoría de Millikan para los estados y procesos mentales, tratando al mismo tiempo de establecer algunos puntos de comparación con la teoría de Fodor, examinada en la sección anterior.

Contrariamente al planteamiento de Fodor, Millikan 1986 piensa que no hay distinción entre la mente concebida como manipuladora de símbolos (manipulación que es posible en virtud de la forma de los símbolos) —o la mente como una máquina impulsada sintácticamente, como lo plantea Fodor— y la mente concebida como una maquinaria semántica, es decir, como un mecanismo que opera con contenidos. Consecuentemente, para Millikan no es necesaria una doble aproximación teórica a la mente. Recordemos que para Fodor es necesario separar la aproximación computacionalista para los procesos y estados mentales de la aproximación al contenido, que, según hemos visto, tiene un carácter metafísico-ontológico. Para Millikan, cualquier diferencia en la forma de los símbolos que entran en una secuencia de procesamiento debe concordar con diferencias en el contenido de dichos símbolos.

Una teoría naturalista de la mente que no separe la forma del contenido se logra mediante la identificación y definición de funciones propias para los elementos intencionales y mediante la especificación de cómo los elementos en cuestión ejecutan su función propia de acuerdo a una explicación Normal, esto es, de acuerdo a ejecuciones exitosas en el pasado bajo condiciones Normales. Las ejecuciones exitosas de los mecanismos intencionales se reflejan, como en el caso de los mecanismos biológicos, en su continua proliferación en la clase de organismos que los utilizan.

Las preguntas que surgen en este punto son, primero, cómo los deseos y creencias —considerados casi en forma unánime como los elementos intencionales por excelencia— pueden categorizarse biológicamente y, segundo, cómo tal concepción biológica puede exhibir —contrariamente a lo que plantea Fodor— la inseparabilidad de forma y contenido. Con el objeto de responder ambas preguntas, consideraremos, en primer lugar, cómo opera la categorización biológica para el caso de los deseos.

En general, los deseos tienen como función propia ayudar a producir aquello que los satisface. Un deseo es individuado por su contenido. La individuación por el

contenido permite la identificación del objeto que satisfaría tal deseo. Sin embargo, el hecho de que alguien tenga un determinado deseo no es una condición suficiente para la ejecución de la función propia de ese deseo, esto es, ayudar a producir el objeto que lo satisfaría. Hay ciertas condiciones Normales que deben ser satisfechas para que el deseo en cuestión pueda ejecutar su función propia de manera exitosa. Supongamos, por ejemplo, que Juan desea comer una manzana. Para que ese deseo pueda ejecutar su función propia hay varias condiciones que deben satisfacerse. Entre muchas otras, podemos mencionar a modo de ilustración: la creencia de Juan de que las manzanas son saludables, nutritivas y que saben bien; la creencia de Juan de que no es difícil obtener manzanas; en la dimensión objetiva, el hecho de que no hay grandes obstáculos para obtener manzanas (por ejemplo, si Juan estuviera en medio del desierto de Sahara, entonces los obstáculos factuales harían altamente improbable la satisfacción de su deseo). El ejemplo muestra que hay al menos dos grupos de condiciones que debieran ser satisfechas. Un primer grupo de acuerdo al cual el deseo individuado entra en una red de relaciones inferenciales con otros estados mentales. En virtud de su participación en esta malla de relaciones inferenciales el deseo mantiene su plausibilidad, lo que evita la posibilidad de que se desvanezca. El segundo grupo de condiciones expresa las situaciones ambientales que, si se dieran, podrían facilitar la satisfacción del deseo. Quizás el deseo no sea realmente satisfecho debido a que alguna de las condiciones Normales que hemos enumerado más arriba no se ha cumplido. Sin embargo, al individuar el deseo por su contenido, se ha explicitado su función propia y al menos se han esbozado las condiciones Normales cuya satisfacción permite la ejecución de la función propia del deseo en cuestión.

Es importante enfatizar en este punto que es esencial para la especificación de la función propia de un deseo el que éste sea individuado por su contenido. Este modo de individuación del deseo determina su rol en los procesos inferenciales que son parte de las condiciones Normales para la ejecución de su función propia. Lo anterior establece una diferencia crucial con respecto a la teoría computacional de los procesos y estados mentales, de clara orientación funcionalista, sostenida por Fodor. En la teoría de Fodor, los deseos y los estados mentales en general son individuados por su rol funcional en una secuencia de procesamiento; y, para estos efectos, el contenido no es tomado en consideración.

Con respecto a su individuación, las creencias exhiben grandes diferencias respecto de los deseos. Para Millikan, las creencias son individuadas por su función propia y por la ejecución de esa función propia en condiciones Normales. Recordemos que en el caso de los deseos la individuación del contenido de un deseo era suficiente para identificar su función propia. En general, puede decirse que la función propia de una creencia consiste en participar en los procesos inferenciales que contribuyen a la satisfacción de los deseos o que ayudan a sustentar otras creencias. Pero, para que esta función propia sea ejecutada de una manera adecuada, la creencia en cuestión debe satisfacer una condición de verdad. Si así no fuera, la cadena inferencial en que la creencia eventualmente participaría sería defectuosa. La satisfacción de una condición de verdad es para Millikan una condición Normal para la ejecución de la función propia de una creencia. Por ejemplo, cuando Juan cree que hay manzanas en el refrigerador su creencia tiene como condición Normal —en este caso una condición de verdad— que, de hecho, haya manzanas en el refrigerador. Cuando la creencia de Juan es individuada tomando en

consideración la satisfacción de esta condición Normal (i.e. condición de verdad), entonces la creencia puede ejecutar su función propia de participar en cadenas de inferencias de manera adecuada. Si así no fuera, la creencia tendría que ser considerada como *aNormal*, esto es, como un estado mental que no desempeña su función ajustándose a una explicación Normal. De acuerdo a lo dicho, el contenido de una creencia y la satisfacción de su condición de verdad son los factores cruciales en el éxito y proliferación de los procesos inferenciales. Como en el caso de la caracterización de los deseos, las diferencias con la aproximación de Fodor, quien, en opinión de Millikan, separa la forma del contenido de los estados mentales, pueden, también en este caso, percibirse claramente.

Lo que hemos expuesto más arriba corresponde a la teoría que Millikan sostiene para los procesos y estados mentales. La caracterización de los deseos y creencias en términos de categorías biológicas, esto es, en términos de funciones propias ejecutadas en condiciones Normales y de acuerdo a una explicación Normal, no parece ser una reducción total a la biología. Sin embargo, Millikan (1986:77-78) piensa que tal caracterización de los procesos y estados mentales puede ser útil para el neuropsicólogo con el objeto de que pueda identificar las estructuras así caracterizadas en el cerebro. Quizás sería posible imaginar alguna manera de identificar los mecanismos de fijación de la creencia y los mecanismos que conectan los deseos con la acción desde un punto de vista neurofisiológico. No obstante, pensamos que sería muy difícil para el neuropsicólogo trabajar con las caracterizaciones de Millikan, ya que contienen demasiadas nociones semánticas e intencionales. Lo anterior puede verse en forma más clara si se considera que los deseos son individuados por sus contenidos y que dicha individuación expresa el objeto intencional de esos deseos. En el caso de las creencias, lo que se requiere como condición Normal es que la creencia sea individuada por su contenido y que las condiciones de verdad sean satisfechas. Con caracterizaciones de los mecanismos intencionales como las expuestas, el neuropsicólogo tendría la difícil tarea de identificar estructuras neurofisiológicas teniendo contenidos como punto de partida. Esto hace su tarea extremadamente difícil, por no decir imposible. Al parecer, entonces, lo que tiene prioridad, con el objeto de hacer la tarea de identificar estructuras neurofisiológicas con mecanismos intencionales más plausible —si es que realmente es una tarea plausible—, es expresar la noción de contenido en términos biológicos. Para estos efectos, Millikan propone lo que ella llama una teoría biológica del contenido o biosemántica (véase Millikan 1989). Desafortunadamente, dicha teoría no propone una noción biológicamente genuina de contenido. Como la teoría que Millikan sostiene para los procesos y estados mentales, su teoría biológica del contenido exhibe una gran cantidad de nociones semánticas e intencionales, y esto hace aun más difícil la tarea que ella visualiza para el neuropsicólogo.

En lo que sigue examinaremos la teoría del contenido que propone Millikan, tratando, al mismo tiempo, de diferenciar su teoría de la propuesta por Fodor.

2.2. Biosemántica

Fodor piensa que las teorías teleológicas y evolucionistas del contenido son teorías causales. Para el autor mencionado, este tipo de teorías trata de especificar contextos causales Normales e idealizados en los cuales el contenido de una representación es

determinado. Además, Fodor ve este tipo de teorías como una solución causalista al problema de la disyunción y, en su opinión, una solución errónea. Sin embargo, la teoría del contenido de Millikan no tiene, contrariamente a lo que piensa Fodor, una orientación causalista. Ciertas consideraciones causalistas tienen lugar en algunos puntos de su teoría, pero para ella la perspectiva causal no es esencial para dar una solución al problema de la determinación del contenido intencional. Millikan piensa que el problema del contenido consiste en establecer las condiciones bajo las cuales una representación corresponde con eventos en el mundo. Para ella, una representación es un signo producido por un sistema en su interacción con factores ambientales. Pero, en su opinión, si tratamos el problema del contenido de las representaciones solamente teniendo en cuenta los factores productivos o causales de las representaciones, no obtendremos una solución al problema en consideración. Si nos centramos solamente en el examen de los factores causales que hicieron posible la producción de un signo, el problema de la intencionalidad del contenido, a juicio de Millikan, quedaría sin tocar. Un sistema puede producir signos que portan información acerca de los factores ambientales que actuaron como causas de dichos signos; pero, de la consideración de estos factores causales e informacionales no puede concluirse que los signos así producidos tienen la función específica de representar unívocamente algo en el mundo. Si la función representadora aparece como algo accidental, cuando se la considera desde el punto de vista de los factores causales que intervienen en la producción de representaciones por parte de un sistema, entonces no hay manera de explicar la determinación del contenido mediante la especificación de los contextos causales en los cuales se producen dichas representaciones.

Para Millikan, la función propia de una representación consiste en que su contenido está en una relación de correspondencia con eventos en el mundo. Las condiciones para la determinación del contenido intencional, y también el establecimiento de la función propia de una representación, solamente pueden darse, a juicio de Millikan, si el problema del contenido es considerado mediante el examen del consumo, y no de la producción, de representaciones. Ella expresa este punto como sigue:

El problema para el naturalista es describir la intencionalidad... [y] el problema de la producción de representaciones no le concierne. A pesar de que una representación siempre es algo que es producido por un sistema cuya función propia es hacer que esa representación corresponda mediante reglas con el mundo, lo que la regla de correspondencia es, lo que define su función, está determinado completamente por los consumidores de representaciones. (1989:286)

Millikan piensa que si se trata el problema del contenido considerando la parte de un sistema que consume o interpreta representaciones será posible no sólo establecer las condiciones de determinación del contenido, sino también definir la función propia de una representación. A su juicio, al usar esta estrategia argumentativa podrá evitar el problema de la disyunción, el cual las teorías teleológicas que enfatizan la parte productora de representaciones no pueden enfrentar adecuadamente. Por otra parte, Millikan piensa que este modo de aproximación al problema del contenido le permitirá evitar lo que ella considera como la consecuencia paradójica de la doble aproximación exhibida por la teoría de las representaciones de Fodor. Según ella, esta doble aproxi-

mación implica separar el problema de la definición del rol representacional del problema de la determinación del contenido de una representación.

Con respecto al último de los problemas enunciados, Millikan piensa que la función propia de una representación es la de establecer una relación de correspondencia con aquello que representa. La relación de correspondencia se establece mediante reglas que hacen posible la proyección del signo representacional sobre eventos que acaecen en el mundo. Las reglas de correspondencia están determinadas por el mecanismo interpretativo del organismo en el cual el signo se produce. Como la determinación de las reglas de proyección es, en cierto sentido, la especificación de las condiciones que permiten alcanzar la relación de correspondencia para un determinado signo representacional, entonces parece a primera vista claro que el problema de la determinación del contenido no puede separarse del problema de la definición de la función propia de una representación. Millikan critica a Fodor por presentar una doble teoría de las representaciones: una, donde el rol causal de una representación está determinado por su rol funcional en una cadena de computaciones y, la otra, donde el contenido de una representación aparece determinado causalmente. En opinión de Millikan, la consecuencia paradójica de la teoría doble de Fodor es que, al separar de este modo el examen de la noción de representación, un signo mental aparece como desempeñando el rol de una representación meramente por su rol funcional, sin ser una representación de nada en particular. Como ya se dijo, Millikan cree que el modo de evitar esta consecuencia, aparentemente paradójica, es mostrar que los mecanismos que consumen o interpretan signos determinan el contenido de una representación y también su función propia (o, lo que aquí hemos llamado también, su rol representacional).

Con respecto al problema de la disyunción, Millikan afirma que dicho problema surge porque algunas teorías teleológicas centran su atención en el examen de aquella parte de un organismo que produce representaciones. Dichas teorías toman en consideración las relaciones causales entre la representación y su representado en circunstancias ideales. El producto de esa relación causal es un signo que porta información acerca de aquellos factores que lo produjeron. Sin embargo, los factores que contribuyen a la producción de un signo son múltiples; en consecuencia, la información que porta el signo producido refleja tal multiplicidad. De modo que al asumir la tarea de establecer las condiciones para la determinación del contenido sobre bases causales (es decir, sobre la base de la información que el signo producido porta acerca de aquello que lo causó), el resultado inevitable serán asignaciones disyuntivas de contenido. Para Millikan, una solución plausible al problema de la disyunción desde una perspectiva teleológica consiste en mostrar cómo la parte que consume o interpreta representaciones en un organismo determinado puede usar la información contenida en un signo entregado por la parte productora para ejecutar tareas que son beneficiosas para la supervivencia de ese organismo. Al hacer esto el teórico de orientación teleológica será capaz, en opinión de Millikan, de identificar el objeto intencional de una representación y, al mismo tiempo, determinar unívocamente el contenido de esa representación.

Para ilustrar su solución al problema de la disyunción, Millikan usa el ejemplo, analizado por Dretske 1986 y 1988, de los magnetosomas de las bacterias de mar. Los magnetosomas son mecanismos que, al ser activados, impulsan a las bacterias del hemisferio norte hacia el norte geomagnético. El hecho de que las bacterias se dirijan en esa dirección es vital para su supervivencia, ya que pueden encontrar agua libre de

oxígeno. Los magnetosomas actúan como mecanismos indicadores que producen un tipo de signo que porta información acerca de los hechos que han posibilitado su producción (en este caso la dirección del norte geomagnético). Pero, teniendo presente el uso que el organismo hace de la información para ejecutar tareas que son vitales para su supervivencia, habría que decir que lo que el signo representa en el caso de las bacterias de mar es la dirección en que se encuentra el agua libre de oxígeno. Si se dijera que el tipo de signo en cuestión representa aquellos factores causales que contribuyeron a su producción (esto es, la dirección del norte geomagnético), el comportamiento de la bacteria, orientado a la dirección en la que se encuentra el agua libre de oxígeno, podría concebirse como algo accidental y no sería posible describirlo de acuerdo a lo que Millikan llama una explicación Normal. Consideraciones similares pueden hacerse en torno al ya extensamente analizado ejemplo del detector de insectos voladores de la rana y del consecuente reflejo que permite al animal atrapar su presa. La parte productora de la rana produce una imagen retinal como respuesta a puntos negros que aparecen en su entorno. Sin embargo, de acuerdo a la visión de Millikan, no puede decirse que los movimientos reflejos que dicha imagen retinal activa estén dirigidos a puntos negros, puesto que bajo esa descripción esos movimientos no exhibirían ventaja adaptativa alguna. Por lo tanto, decir que las imágenes retinales de la rana representan insectos voladores ingeribles —para abreviar, moscas— está en concordancia con una explicación Normal de la función propia del mecanismo que la rana posee para detectar moscas.

Es obvio que en el caso de las ranas y de las bacterias de mar no puede decirse que dichos organismos interpretan los signos producidos de acuerdo a ciertos conceptos. El contenido de la representación es asignado externamente en concordancia con lo que Millikan llama una explicación Normal. Como se dijo, una explicación Normal contiene las condiciones históricas bajo las cuales las ejecuciones de un mecanismo fueron exitosas y debe también contener referencias a los beneficios adaptativos que un organismo obtiene al usar un cierto mecanismo que desempeña una determinada función.

La argumentación de Millikan para evitar el problema de la disyunción parece operar en forma adecuada para sistemas simples como las ranas y las bacterias de mar. Sin embargo, es difícil ver cómo puede operar en sistemas más complejos como los humanos, los cuales poseen representaciones cuyo propósito biológico parece difícil de determinar con exactitud. Podría decirse que la complejidad de la parte que consume representaciones en un organismo humano está dada por la variedad de usos que dichos organismos pueden dar a un mismo tipo de signo proporcionado por la parte productora. En otras palabras, un signo entregado por la parte productora de representaciones de un organismo humano puede ser usado por el mecanismo interpretador para ejecutar, en forma simultánea, varias tareas que son consideradas como beneficiosas para ese organismo. Si esto es así, entonces la asignación de funciones para una representación no identificaría unívocamente su objeto intencional. Ahora bien, si la asignación de funciones no hace posible la identificación unívoca del objeto intencional, tenemos nuevamente el problema de la disyunción, ahora bajo la forma de indeterminación de la función propia de una representación.

Por otra parte, alguien puede objetar que los casos de la rana y de la bacteria de mar no son casos genuinos de contenido intencional, puesto que es extremadamente difícil

explicar cómo estos sistemas simples tienen representaciones fallidas. Si ponemos a un grupo de bacterias del hemisferio norte en el hemisferio sur, morirán en forma inevitable puesto que la información que sus magnetosomas les entreguen las impulsarán en esta nueva situación hacia aguas que les son letales; y si ponemos a una rana en un laboratorio en el cual hay un dispositivo especial que proyecta puntos negros (véase Dretske 1988), la rana morirá de hambre al ser su reflejo activado por puntos negros que no satisfacen sus necesidades alimenticias. Estos dos casos muestran que la parte productora y la consumidora de representaciones están operando en forma adecuada—esto es, están representando los eventos del mundo en el modo que se supone deben representarlos—, pero, al mismo tiempo, están fracasando al representar una condición del mundo que es vital para su supervivencia. A esto Millikan respondería que se trata de casos genuinos de representación fallida puesto que hay algunas condiciones Normales que no han sido satisfechas para la adecuada ejecución de la función propia de las representaciones en cuestión. En los casos examinados, ambos tipos de organismos han sido sacados del entorno ambiental en el que los mecanismos representativos desempeñaban adecuadamente la función que les era propia. En consecuencia, las condiciones aNormales, para Millikan, explican los casos de representación fallida. Sin embargo, los dos casos mencionados provocan dudas acerca del modo en que puede determinarse el objeto intencional de una representación siguiendo la propuesta de Millikan. Las respuestas de la rana y de las bacterias de mar bajo condiciones Normales son las mismas que estos organismos exhiben bajo condiciones similares pero aNormales. Siendo esto así, parece difícil establecer si acaso el objeto intencional de las respectivas representaciones es el modo en que los organismos representan una determinada condición en el mundo (puntos negros y norte geomagnético) o esa condición misma (moscas y agua libre de oxígeno). Ahora bien, si no hay claridad respecto de la determinación del objeto intencional, tampoco habrá suficiente claridad para explicar los casos de representación fallida. En los casos que hemos examinado no es posible decidir si los organismos en cuestión tienen representaciones fallidas o están representando los estados del mundo del modo en que se supone deben representarlos. Volveremos a considerar estos ejemplos en conexión con el problema de la representación fallida cuando examinemos la teoría de Dretske en la tercera sección de este trabajo.

2.3. *Signos portadores de información e iconos intencionales*

En esta última parte de esta segunda sección trataremos de mostrar hasta qué punto la teoría de Millikan es contraria a una concepción causal e informacional del contenido. Lo que aquí digamos puede ser útil para diferenciar la aproximación de Millikan del tipo de semántica que está en la base de la concepción de Fodor. Las consideraciones que aquí se hagan serán, al mismo tiempo, útiles para establecer hasta qué punto el proyecto de Millikan exhibe una orientación naturalista.

En su libro *Language, thought and other biological categories* (1984: esp. Par II), Millikan está principalmente interesada en lo que ella considera un caso paradigmático de signos a los que llama iconos intencionales. Para estar incluido en el paradigma de los iconos intencionales, un signo debe cumplir la condición de ser un elemento que está a medio camino entre un mecanismo productor y un mecanismo interpretativo. Estos dos mecanismos están en una relación cooperativa tal que la presencia y acción de cada

uno de ellos es un factor que cuenta para el adecuado desempeño del otro. Un icono intencional sirve para adaptar el mecanismo interpretativo a condiciones tales que el mecanismo en cuestión pueda ejecutar su función propia. La función propia del mecanismo interpretativo es, para expresarlo en términos muy generales, hacer posible la correspondencia entre el icono intencional y los eventos en el mundo que éste representa mediante reglas de proyección. Las representaciones son un tipo especial de iconos intencionales que involucran la identificación de los referentes sobre los cuales los constituyentes de la representación se proyectan. Millikan piensa que los signos naturales, caracterizados como portadores de información, no pueden considerarse por sí mismos como iconos intencionales y, por lo tanto, no son representaciones. Esto es así puesto que su producción no presupone la existencia de un mecanismo interpretativo que el signo en cuestión tenga que adaptar. Siendo así, la correspondencia no es algo que esté contenido dentro de las funciones de este tipo de signos. Los signos naturales pueden ser usados por un intérprete de un modo análogo al que son usados los iconos intencionales, pero esto no significa que los signos naturales tengan por sí mismos la función de adaptar un mecanismo interpretativo a condiciones que harán posible la consecución de la correspondencia entre el signo y lo significado. El uso derivativo que un intérprete haga de un signo natural no estará en concordancia con una explicación Normal (como es el caso con los iconos intencionales), sino más bien concordará con leyes naturales. Millikan caracteriza el uso derivativo de los signos naturales de la siguiente manera:

Un signo natural puede ser usado por un intérprete como si fuera un icono intencional indicativo que proyecta algo en concordancia con leyes naturales operando in situ. Este tipo de signos hace esto no en forma Normal, en el sentido médico o biológico. Más bien debe hacer esto en forma relativamente confiable o normal, en el sentido estadístico. Esto es, el tipo de situación en la cual, o las condiciones bajo las cuales, el signo es determinado por leyes de la naturaleza para proyectarse sobre aquello que significa, debe ser una situación o condiciones que son estadísticamente normales. (1984:118-119)

Como se dijo, los signos naturales son iconos intencionales solamente en un sentido derivativo. No pueden ser falsos, ya que no pueden ejecutar su función de un modo aNormal. Pero pueden ser falsos derivativamente, es decir, por referencia a un intérprete que los usa como si fueran signos. Estas razones llevan a la conclusión de que los signos naturales (como el humo que indica fuego o las nubes negras que indican una alta probabilidad de lluvia) pueden tener solamente una intencionalidad derivada, esto es, derivada de los propósitos y del uso que un intérprete hace de ellos.

De acuerdo a lo dicho, los signos naturales (o signos que son el output de un mecanismo productivo y que no tienen entre sus funciones la de cooperar adaptativamente con un mecanismo interpretativo, o signos que meramente portan información acerca de los eventos que han contribuido a causarlos) no pueden tener como su función propia el corresponder con algo en el mundo externo por medio de una regla de proyección. Si esto es así, ningún contenido puede asignarse unívocamente a ellos. Solamente los signos que sirven de mediadores entre un mecanismo productivo y un mecanismo interpretativo y que tengan, además, como función propia el adaptar el mecanismo interpretativo a las condiciones bajo las cuales esos signos se proyectan en el mundo, pueden desempeñar la función propia de una representación y tener, por lo

tanto, un contenido determinado. Según esta visión sumaria de la concepción de Millikan acerca de los signos naturales en tanto diferentes de los iconos intencionales, una teoría del contenido como la que sostiene Fodor no tiene cabida. Las conexiones causales, desde el punto de vista de Millikan, solamente pueden producir signos naturales, y los signos naturales no pueden tener de manera intrínseca un rol representacional y un contenido determinado.

Hemos mostrado que la aproximación naturalista de Millikan deja fuera, como un aspecto no fundamental para una teoría del contenido, la dimensión causal. Para ella, lo esencial en la determinación del contenido de una representación es el rol que le cabe desempeñar al mecanismo consumidor o interpretativo de un signo al adaptarlo a las condiciones bajo las cuales sus elementos se proyectan en el mundo. De acuerdo a lo anterior, el tipo de semántica que Millikan está proponiendo no es una semántica de los indicadores —para usar la terminología de Block 1986—, la cual agrupa a todas las semánticas causales e informacionales, incluida la de Fodor, sino más bien una semántica de los usos que un organismo hace de la información contenida en un determinado signo. Podríamos llamar a este tipo de semántica una semántica de los mecanismos interpretativos. Lo que tendríamos que requerir de Millikan, teniendo presente que lo que estamos buscando son las bases de una semántica naturalizada, es una noción naturalista de un mecanismo interpretativo. Lamentablemente, no hemos podido encontrar en Millikan un intento completo de naturalizar dicha noción. Su solución al problema del contenido, fundada en un examen de los mecanismos que consumen o interpretan signos, contiene demasiadas nociones intencionales. La categorización de elementos intencionales en términos de funciones propias, condiciones Normales y explicaciones Normales no es, al parecer, suficiente, especialmente si la especificación de esas categorizaciones se hace en términos intencionales. A modo de ilustración, podemos decir que, en su artículo de 1989 acerca del contenido representacional, ella especifica las mismas condiciones Normales para la ejecución de la función propia de una representación en el contexto de las creencias que la especificación que formulara en su artículo de 1986 para la ejecución de la función propia de una creencia involucrada en el procesamiento mental; esto es, la individuación mediante el contenido y la satisfacción de una condición de verdad. Lo anterior pone de manifiesto que en su teoría biosemántica, al igual que en su teoría de los procesos y estados mentales, el contenido no es naturalizado sino presupuesto. Por otra parte, suponiendo que desde su perspectiva pudiera naturalizarse el contenido, ello requeriría previamente naturalizar la noción de intérprete o usuario de una representación.

De acuerdo a lo expresado, la conclusión de nuestro examen de la teoría de Millikan, y teniendo siempre presente que nuestra principal preocupación es un problema de la naturalización del contenido, es que su propuesta no constituye una teoría completamente naturalista del significado. A pesar de que ella consigue llevar adelante la difícil tarea de categorizar los mecanismos intencionales en términos biológicos, las especificaciones requeridas por aquellas categorizaciones contienen demasiados elementos conceptuales de carácter intencional.

En la última sección de este trabajo examinaremos en forma breve la teoría del contenido propuesta por Dretske. Nos parece que Dretske es capaz de conciliar la semántica de los indicadores, que está a la base de las teorías causales e informacionales, de las cuales la de Fodor es su ejemplo más refinado, con una semántica de los

mecanismos interpretativos, como la que, según vimos, sostiene, con algunas dificultades, Millikan. Como veremos, su aproximación naturalista difiere del tipo de enfoque reductivo y ontológico sostenido por Fodor y del intento de expresar los ítemes intencionales en términos de categorías biológicas sostenido por Millikan.

3. DRETSKE Y EL ROL CAUSAL DEL CONTENIDO EN LA PRODUCCIÓN DEL COMPORTAMIENTO

El último libro de Dretske, *Explaining behavior* (1988), tiene como propósito central entregar una visión del rol que le cabe al contenido en la explicación causal del comportamiento. Según este filósofo, el hecho de mostrar el rol del significado en la producción del comportamiento permitirá comprender mejor el papel que le cabe a las razones (i.e. creencias y deseos) en un mundo regido por causas. Sin embargo, nuestro interés no está centrado en aspectos del tema general del libro, sino más bien en un tópico bien específico que encontramos en el tratamiento que Dretske hace de lo que él llama 'sistemas representacionales' (véase Dretske 1988, cap. 3). Al examinar su teoría de los sistemas representacionales queremos mostrar que ésta tiene tres virtudes: primero, que hace posible la conciliación del enfoque informacional y causal con lo que aquí hemos llamado una semántica de los mecanismos interpretativos; segundo, que dicha teoría indica cómo una semántica de los mecanismos interpretativos puede naturalizarse; y, tercero, que la teoría de Dretske muestra que se puede entregar una visión naturalista del contenido sin necesidad de una maniobra reductiva. Estas conclusiones parecen a primera vista optimistas respecto de una eventual solución al problema que estamos examinando, puesto que son aún provisionales y, como tales, necesitan todavía mayor argumentación y fundamento. No obstante, señala una ruta que puede ser promisoría para solucionar el problema de formular una semántica naturalista.

La concepción de las representaciones sostenida por Dretske 1988 está centrada en lo que él llama sistemas representacionales (SR). Los elementos de los SR están en una relación indicadora con aquello que representan. Dretske caracteriza un sistema representacional como aquel:

...cuya función es la de indicar cómo están las cosas respecto de algún otro objeto, condición o magnitud. Si la función del SR es indicar si acaso O [objeto] está en la condición de A o B, por ejemplo, y el modo en que SR ejecuta esta función (cuando la ejecuta) es ocupando uno de dos estados posibles, a (que indica que O es A) y b (que indica que O es B), entonces a y b son los elementos expresivos de SR, y lo que ellos representan (acerca de O) es que es A (en el caso de a) y que es B (en el caso de b). (1988:52)

Dretske distingue tres tipos de sistemas representacionales, de acuerdo a las características que exhibe la relación indicadora entre los elementos expresivos de un SR y lo que ellos representan. El primer tipo de sistema que Dretske distingue es aquel cuyos elementos expresivos tienen una relación convencional con sus representados. En este caso, la función indicadora del elemento expresivo que pertenece a ese sistema es asignada y extrínseca al sistema en cuestión. Por ejemplo, un ladrón acuerda con un compañero de oficio que cuando alguien se acerque, mientras uno de ellos está haciendo su trabajo, el otro debe emitir un silbido con tales y tales características, como una señal de alerta.

El segundo tipo de SR contiene elementos expresivos a los cuales, pese a ser signos

naturales en lo que a su etiología se refiere, se les asigna una función indicadora de manera extrínseca y, por lo tanto, convencional. Contrariamente a lo que sucedía con los SR convencionales, la función indicadora que se asigna a este tipo de signos naturales forma parte de la información (en el sentido etiológico) que el signo en cuestión porta. Esa parte específica de la información es seleccionada por nosotros con el objeto de asignar una función indicadora a dicho signo. Como se ha dicho en varios pasajes de este trabajo, los signos naturales contienen una pluralidad de información, puesto que pueden entrar en una pluralidad de correlaciones causales legales. Sin embargo, dado que alguien tiene un propósito especial que guiará la utilización que haga de ese signo natural, esa persona puede seleccionar una correlación legal determinada, asignando de este modo una función indicadora específica a ese signo. Los elementos de este tipo de SR adquieren, entonces, la función de indicar algo específico en virtud del interés que tiene la persona que pretende usarlos. Un ejemplo de este tipo de signos son los anillos que encontramos en los troncos de los árboles. Si un especialista está interesado en medir la edad de los árboles, ese especialista asignará a los anillos, basado en una correlación legal, dicha función indicadora. Pero si otro especialista está interesado en buscar indicios para establecer las características de los ciclos climáticos secos y lluviosos en una determinada región, entonces, y también basados en ciertas correlaciones legales, asignará a los anillos esa función indicadora específica. En resumen, en este segundo tipo de SR encontramos signos producidos de acuerdo a correlaciones causales legales (i.e. leyes naturales), pero con una función indicadora asignada extrínsecamente y seleccionada, en concordancia con los intereses de un usuario, de un rango de correlaciones legales de las cuales el signo en cuestión puede formar parte.

El segundo tipo de SR no posee, sin embargo, uno de los rasgos cruciales de los casos genuinos de representación. Los signos que constituyen este tipo de sistema representacional son incapaces de representar fallidamente, y éste, a juicio de Dretske, es uno de los rasgos esenciales de los signos que exhiben una intencionalidad intrínseca. A modo de ilustración, tomemos el caso de un niño que tiene pintas rojas en la cara. Al ver las pintas, el médico diagnostica sarampión. Pero supongamos que de hecho el niño no tiene sarampión. ¿Podemos decir que las pintas rojas en la cara del niño representan en forma fallida el sarampión? La respuesta es obviamente negativa. La explicación de Dretske para esta respuesta negativa es que las pintas rojas no tienen como su función indicadora específica —si es que realmente tienen alguna— el representar sarampión. Es el médico, como un usuario o intérprete del signo, el que ha seleccionado la correlación legal errónea. En este caso, la correlación legal que conecta las pintas rojas con el sarampión simplemente no es aplicable. Lo anterior muestra que el tipo de signo en consideración no puede indicar o representar fallidamente; en otras palabras, no puede ser falso (compárese esta visión con la teoría de los signos naturales de Millikan, mencionada en 2.3). Casos como los mencionados más arriba son, en opinión de Dretske, casos de signos que tienen lo que Grice 1957 llama significado natural; esto es, no pueden significar P sin que P sea el caso.

Pero Dretske está interesado en casos de SR que exhiban intencionalidad genuina; es decir, en aquellos sistemas cuyos elementos exhiban la capacidad de representar fallidamente o falsamente. En otras palabras, Dretske está buscando sistemas representacionales cuyos signos posean lo que Grice 1957 llama 'significado no-natural'; esto es, signos que puedan significar P sin que P sea el caso. Asumiendo una perspectiva

naturalista, Dretske concibe este tipo de SR como compuesto por signos que tienen una función indicadora no-asignada e intrínseca al sistema mismo. Si un sistema satisface estos requerimientos —significado griceano no-natural, una función indicadora específica y no asignada extrínsecamente y, por lo tanto, natural—, entonces dicho sistema puede contar como uno que exhibe intencionalidad intrínseca.

Sin embargo, hay algunos sistemas representacionales simples, como el detector de moscas de la rana y los magnetosomas de las bacterias de mar, que hemos considerado en la sección anterior, que satisfacen esos requisitos pero para los cuales los casos de representación fallida no pueden explicarse en forma clara. Siendo esto así, su clasificación como casos de representación genuina es dudosa. Dretske piensa que los magnetosomas de las bacterias de mar son mecanismos que ejecutan la función indicadora que esa especie de organismos ha desarrollado durante el proceso de evolución. La función indicadora, según esto, forma parte de la fisiología de estos organismos y es heredada genéticamente de sus ancestros. Si el organismo no tiene problemas internos de mal funcionamiento (heredados o debido a algún factor accidental), entonces, dado un cierto estímulo en el ambiente adecuado, el magnetosoma reaccionará de un modo preprogramado produciendo una representación o signo, que, a su vez, causará un movimiento, también programado genéticamente dirigido hacia aguas libres de oxígeno. De acuerdo a la caracterización esbozada, tenemos que asumir que el mecanismo en cuestión está calibrado para operar adecuadamente bajo ciertas condiciones y en un ambiente del tipo en el que la función indicadora se desarrolló durante el proceso de evolución. Dadas estas circunstancias, el sistema en cuestión producirá la representación correcta. Si esto fuera así, tendríamos que decir, con Millikan, que los casos de representación fallida ocurren cuando no se da alguna de las circunstancias normales que hemos enunciado.

No obstante, como lo señala Dretske, en sistemas simples como los que estamos considerando, los casos de representación fallida son todavía problemáticos y difíciles de explicar. Si ponemos a una rana en un laboratorio —para usar el ejemplo de Dretske— en el cual una gran cantidad de pequeñas sombras son proyectadas por un mecanismo especial para simular moscas, el detector de moscas de la rana será estimulado y las representaciones que este mecanismo produzca activarán el reflejo que permite al animal aprehender su presa. Aparentemente, el detector de moscas de la rana está operando adecuadamente al reaccionar ante puntos oscuros móviles y produciendo los indicadores que gatillan los movimientos reflejos de la lengua que permiten la aprehensión del insecto. Sin embargo, las condiciones ambientales en las cuales el mecanismo está operando no son aquellas en las cuales dicho mecanismo usualmente opera en forma satisfactoria. Como lo señalamos en su oportunidad, Millikan afirmaría, sin dejar lugar a ninguna duda, que éste es un caso de representación fallida puesto que hay ciertas condiciones ambientales Normales que no se han dado. Pero el problema de Dretske es si acaso puede determinarse que el sistema representacional de la rana está produciendo una representación fallida, y no si éste es un caso de representación fallida porque no se han satisfecho ciertas condiciones ambientales. Dretske piensa que el problema de si acaso estos sistemas simples pueden producir representaciones falsas es de difícil solución. Lo que se produce en estos casos, en opinión de Dretske, es una indeterminación de la función indicadora del detector, y es esta indeterminación lo que hace difícil explicar los casos de representación fallida para los sistemas representacio-

nales simples que estamos considerando. Dretske explica estos casos de indeterminación de la función indicadora como sigue:

La representación fallida depende de dos cosas: la condición del mundo que está siendo representada y el modo en que el mundo es representado. La última, como hemos visto, está determinada no por lo que el sistema indica acerca del mundo, sino por lo que éste tiene como función indicar acerca del mundo. Y mientras se mantenga esta indeterminación de la función no hay un sentido claro según el cual la representación fallida pueda ocurrir. Sin una función determinada, uno puede, por así decirlo, siempre eximir de error a un SR, y eliminar, así, la ocurrencia de representaciones fallidas cambiando lo que se supone [que la representación] indica, cambiando lo que es su función indicar. (1988:69)

De acuerdo a la cita, Dretske reconoce el problema de la disyunción bajo la forma de indeterminación de la función indicadora. La rana, cuando es ubicada en el laboratorio, representa fallidamente por la condición del mundo que está siendo representada; pero no puede decirse que está representando erróneamente por el modo en que su detector indica lo que se supone debe indicar acerca del mundo. La prueba de esto último está en el hecho de que las sombras oscuras proyectadas seguirán activando el reflejo de aprehensión de la rana aunque ésta muera de inanición. El problema es qué función indicadora asignar al detector de moscas de la rana. El caso que se ha examinado muestra que la función es difícil de determinar. Lo que la rana representa, si es que algo representa, puede caracterizarse como disyuntivo: o moscas o puntos oscuros móviles. Siendo así, al menos para estos sistemas representacionales, los casos de error o representación fallida no pueden explicarse claramente.

El problema de la indeterminación de la función indicadora puede resolverse, en opinión de Dretske, no en el nivel de sistemas simples como las ranas y las bacterias de mar, sino en el nivel de los organismos que son capaces de aprender. Dretske parte del supuesto de que estos últimos tienen varios detectores que indican o transmiten información acerca del mundo. Si lo que alguno de estos detectores indica acerca del mundo se conecta a mecanismos que controlan y producen el comportamiento, el detector adquiere una función indicadora específica. La función indicadora adquiere especificidad debido a que los organismos en cuestión aprenden, sobre la base de un esquema de recompensa y castigo, a comportarse de determinada manera en virtud de lo que el indicador indica. De acuerdo a esta explicación, el indicador sirve como un mapa que ayuda a guiar y a controlar el comportamiento de esos organismos. Si esto es así, el éxito de un cierto comportamiento aprendido depende de aquello que el indicador indica o representa, y no solamente, como en el caso de la rana y de las bacterias de mar, del modo en que el indicador indica lo que se supone debe indicar. Cuando los indicadores o representaciones adquieren de este modo su función, el contenido pasa a tener un rol en la explicación causal del comportamiento. El que el contenido tenga un rol en la explicación de cómo se produce el comportamiento no significa que el contenido por sí solo sea una causa del comportamiento, sino que, como se señaló, es un factor que tiene un rol importante en la producción del mismo.

Obviamente, hay mucho más que decir acerca de la teoría de Dretske y acerca de las implicancias que ésta tiene para el problema de la naturalización del contenido. Pero dejaremos un tratamiento más profundo de estas cuestiones para otra ocasión. Por ahora, y basados en la visión sumaria que hemos entregado de la teoría de Dretske, nos

interesa desprender algunas conclusiones respecto de los dos primeros intentos de naturalización del contenido que hemos examinado en forma más extensa en las dos primeras secciones. Las conclusiones en cuestión fueron anunciadas en la Introducción y también al comenzar esta tercera sección. Lo que haremos ahora es explicitarlas un poco más teniendo presente la visión sumaria de la teoría de Dretske esbozada en los párrafos anteriores.

Con respecto a la primera de las conclusiones que anunciáramos al comenzar la tercera sección, es claro que Dretske trata de conciliar dos tipos de aproximación al problema del contenido que para Fodor y Millikan estaban en conflicto: una semántica informacional y una semántica de los mecanismos interpretativos. Dretske no rechaza su anterior teoría semántica informacional, que presentara en *Knowledge and the flow of information* (1981). De hecho, el funcionamiento de los sistemas representacionales está, en último término, fundado en relaciones informacionales. Sin embargo, una teoría informacional del contenido es suficiente para explicar cómo el contenido tiene un rol causal en la explicación del comportamiento de los organismos que tienen representaciones. Para construir una teoría que dé cuenta del rol causal que los estados que tienen un contenido desempeñan en la producción del comportamiento de algunos tipos de organismos, Dretske necesita una semántica de los usos que un organismo hace de la información (lo que aquí hemos llamado, de un modo algo impreciso, una semántica de los mecanismos interpretativos). Al contrario de lo que sucede con Millikan, y esto se conecta con la segunda de las conclusiones anunciadas al comienzo de esta tercera sección, Dretske es capaz de formular una versión naturalista de una semántica de los mecanismos interpretativos mediante el establecimiento del rol del contenido en el control y producción del comportamiento. Con el objeto de establecer dicho rol, Dretske basa su aproximación en los datos de la etología sin perder de vista el carácter filosófico del problema que está examinando. Esta última observación general, y hasta cierto punto vaga, está conectada con la tercera de las conclusiones anunciadas al comenzar esta sección y con la última de las anunciadas en la Introducción a este trabajo. A pesar de que la teoría de Dretske intenta explicar el rol que algunas estructuras que tienen contenido cumplen en la explicación causal del comportamiento y, en este sentido, es incompleta como una teoría global del contenido, la manera en que él aborda el problema desde un punto de vista naturalista nos parece promisoria, principalmente porque no hay un interés compulsivo en llevar a cabo estrategias reductivas para formular una semántica naturalista, como sucede, en nuestra opinión, con las teorías de Fodor y Millikan. Lo que visualizamos en la aproximación naturalista de Dretske es un intento de hacer compatibles las intuiciones filosóficas acerca del contenido con algunas hipótesis relevantes de una disciplina atinente al problema en cuestión, como lo es la etología.

La moraleja de este trabajo es, entonces, que las aproximaciones reductivistas al problema del contenido intencional no resultan demasiado útiles para comprender el rol del contenido representacional en los organismos naturales que lo poseen, y que es quizás más útil, para estos efectos, conciliar y hacer compatibles nuestras intuiciones filosóficas acerca de la intencionalidad con los datos que algunas disciplinas científicas que se ocupan de los problemas ligados a los fenómenos semánticos e intencionales puedan proporcionar. Es posible que el resultado no sea una naturalización completa y

elegante del contenido intencional, como la que persiguen las teorías reductivas, pero al menos se logrará una comprensión más profunda de su rol en el mundo natural.

REFERENCIAS

- BLOCK, N. 1986. Advertisement for a semantics for psychology. *Midwest Studies in Philosophy*, vol. 10: 615-678.
- CUMMINS, R. 1989. *Meaning and mental representation*. Cambridge, Mass.: MIT Press.
- DENNETT, D. 1987. Evolution, error and intentionality. En D. Dennett (ed.), *The intentional stance*. Cambridge, Mass.: MIT Press.
- DRETSKE, F. 1981. *Knowledge and the flow of information*. Oxford: Blackwell.
- DRETSKE, F. 1986. Misrepresentation. En R. Bogdan (ed.), *Belief*. Oxford: Oxford University Press.
- DRETSKE, F. 1988. *Explaining behavior: Reasons in a world of causes*. Cambridge, Mass.: MIT Press.
- FODOR, J.A. 1975. *The language of thought*. New York: Thomas Y. Crowell.
- FODOR, J.A. 1981. Methodological solipsism as a research strategy in cognitive psychology. En *RePresentations*. Cambridge, Mass.: MIT Press.
- FODOR, J.A. 1983. *The modularity of mind*. Cambridge, Mass.: MIT Press.
- FODOR, J.A. 1985. Fodor's guide to mental representation: The intelligent auntie's vade-mecum. *Mind* 94: 77-100.
- FODOR, J.A. 1987. *Psychosemantics: The problem of meaning in the philosophy of mind*. Cambridge, Mass.: MIT Press.
- FODOR, J.A. 1990. Psychosemantics, or: Where do truth conditions come from? En W. Lycan (ed.), *Mind and cognition*. Oxford: Blackwell.
- FODOR, J.A. 1990. *A theory of content and other essays*. Cambridge, Mass.: MIT Press.
- GRICE, H.P. 1957. Meaning. *Philosophical Review* 66: 377:388.
- JACKENDOFF, R. 1983. *Semantics and cognition*. Cambridge, Mass.: MIT Press.
- JACKENDOFF, R. 1987. *Consciousness and the computational mind*. Cambridge, Mass.: MIT Press.
- JOHNSON-LAIRD, P.N. 1983. *Mental models*. Cambridge, Mass.: Harvard University Press.
- JOHNSON-LAIRD, P.N. 1988a. *The computer and the mind*. Cambridge, Mass.: Harvard University Press.
- JOHNSON-LAIRD, P.N. 1988b. How is meaning mentally represented? En U. Eco et al. (eds.), *Meaning and mental representations*. Bloomington, Indiana: Indiana University Press.
- LYCAN, N.A. 1990. *Mind and cognition*. Oxford: Blackwell.
- MILLIKAN, R.G. 1984. *Language, thought and other biological categories: New foundations for realism*. Cambridge, Mass.: MIT Press.
- MILLIKAN, R.G. 1986. Thoughts without laws: Cognitive science with content. *Philosophical Review* 95, 1: 47-80.
- MILLIKAN, R.G. 1989. Biosemantics. *The Journal of Philosophy* 86, 6: 281-297.
- SEARLE, J. 1983. *Intentionality*. Cambridge: Cambridge University Press.
- STILLINGS, N.A., M.H. FEINSTEIN, J.L. GARFIELD, E.L. RISSLAND, D.A. ROSENBAUM, S.E. WEISLER y L. BAKER-WARD. 1987. *Cognitive science: An introduction*. Cambridge, Mass.: MIT Press.